DOCUMENT RESUME

ED 384 990 EA 026 891

TITLE	Issues To Consider in Moving beyond a Minimal Competency High School Graduation Test.
INSTITUTION	Southeastern Regional Vision for Education (SERVE), Tallahassee, FL.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Jul 95
CONTRACT	RP91002010
NOTE	60p.
AVAILABLE FROM	NEFEC/SERVE, Route 1, Box 8500, 3841 Reid Street, Palatka, FL 32177 (Item No. RDMCT; free).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Standards; High Schools; Minimum Competencies; *Minimum Competency Testing; Performance; *State Standards; Student Certification; *Student Evaluation; *Testing Problems
IDENTIFIERS	*Mississippi; *United States (Southeast)

ABSTRACT

This report offers advice on the issues to be considered and the steps to be taken when implementing a high school graduation test. The research was conducted, specifically, to address problems with Mississippi's high school exit test. An external panel, developed by Southeastern Regional Vision for Education (SERVE), reviewed data obtained through a site visit, interviews, and document analysis. Chapter 1 presents introductory information, chapter 2 provides an overview of the Mississippi context, chapter 3 contains an executive summary, and the final chapter contains the full report. The following issues are addressed: curriculum/test specification issues; additional curriculum and instructional considerations; psychometric testing and scoring; education issues; legal issues; policy/administrative issues; and human and financial resource issues. Suggestions are also offered for the sequencing of tasks and using test scores for accreditation purposes. The report contains 65 recommendations, including these: (1) it is legally inappropriate to hold students accountable for passing an assessment that covers materials they have not been taught; (2) multiple-choice items can measure higher order thinking skills and procedures; (3) any "off-the-shelf" test would probably be an unacceptable high school exit test for Mississippi students; (4) requiring any national norm-referencing component of the exit exam poses problems for maintaining curricular validity; (5) the various assessment programs should be closely articulated; and (6) the use of various tests in a performance-based accreditation model requires careful consideration of how to set the performance level and what metric to use. Information for ordering SERVE products is included. (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: EA
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view. TM

ED 384 990

SERVE

Issues to Consider in Moving Beyond a Minimal Competency High School Graduation Test

R&D

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EA 026 891

SouthEastern Regional Vision for Education

Issues to Consider in Moving Beyond a Minimal Competency High School Graduation Test

developed by
An External Panel
Chaired by
Dr. William A. Mehrens

July, 1995

SERVE
SouthEastern Regional Vision for Education

A Special Report Funded By

SERVE
The SouthEastern Regional Vision for Education
Associated with the School of Education
University of North Carolina at Greensboro
P.O. Box 5367
Greensboro, NC 27435
(800) 755-3277

The content of this publication does not necessarily reflect the views or policies of the Office of Educational Research and Improvement, U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This document was produced with funds from the Office of Educational Research and Improvement, U.S. Department of Education, on contract no. RP91002010.

Book Design by
Kelly Killman Dryden

Table of Contents

Chapter 1
Introduction 1

Chapter 2
Background Information for the Report 11

Chapter 3
Executive Summary of the Full Report 15

Chapter 4
The Full Report 19

Issues and Recommendations Regarding the Mississippi Assessment of Academic Proficiency (MAAP) and Other Components of their Assessment for Accountability and Learning Plans

Order Forms
Ordering Information 53

Shipping & Handling Charges 53

Publications Listing 54

Videotape Listing 55

Order Form 57

About the SERVE Laboratory

SERVE, THE SOUTHEASTERN REGIONAL VISION FOR EDUCATION, IS A COALITION OF EDUCATORS, business leaders, governors, and policymakers who are seeking comprehensive and lasting improvement in education in Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina. The name of the Laboratory reflects a commitment to creating a shared vision of the future of education in the Southeast.

The mission of SERVE is to provide leadership, support, and research to assist state and local efforts in improving educational outcomes, especially for at-risk and rural students. Laboratory goals are to address critical issues in the region, work as a catalyst for positive change, serve as a broker of exemplary research and practice, and become an invaluable source of information for individuals working to promote systemic educational improvement.

Collaboration and networking are at the heart of SERVE's mission; the laboratory's structure is itself a model of collaboration. The laboratory has four offices in the region to better serve the needs of state and local education stakeholders. SERVE's Greensboro office manages a variety of research and development projects that meet regional needs for the development of new products, services and information about emerging issues. The development of this manual was funded through such an R&D effort. The laboratory's information office is located in Tallahassee. Field services offices are located in Atlanta, Greensboro, Tallahassee, and on the campus of Delta State University in Cleveland, Mississippi.

To request publications or to join the SERVE mailing list and receive announcements about laboratory publications, contact the SERVE office in Tallahassee (address below).

SERVE-Alabama
50 N. Ripley Street
Gordon Persons Building
Montgomery, AL 36130
334-242-9758
Fax 334-242-9708

SERVE-Florida
345 South Magnolia Drive
Suite D-23
Tallahassee, FL 32301

Lab
904-671-6000
800-352-6001
Fax 904-671-6020

Clearinghouse
800-352-3747

Math Science Consortium
904-671-6033
800-854-0476
Fax 904-671-6010

SERVE-Georgia
41 Marietta Street, NW
Suite 1000
Atlanta, GA 30303
404-577-7737
800-659-3204
SERVE-Line 800-487-7605
Fax 404-577-7812

SERVE-Mississippi
Delta State University
Box 3183
Cleveland, MS 38732
601-846-4384
800-326-4548
Fax 601-846-4402

SERVE-North Carolina
201 Ferguson Building
UNCG Campus
P.O. Box 5367
Greensboro, NC 27435
910-334-3211
800-755-3277
Fax 910-334-3268

SERVE-South Carolina
1429 Senate Street
1008 Rutledge Building
Columbia, SC 29201
803-734-4110
Fax 803-734-3389

CHAPTER 1

Introduction

By Wendy McColskey, Research Program Manager, SERVE

SERVE WORKS WITH STATES, DISTRICTS, and schools to improve educational practices and outcomes. SERVE provides a variety of products and services to the region. Of particular relevance to this report is the work accomplished in the area of student assessment. The lab's work in this area has included a manual for teachers entitled *How to Assess Student Performance in Science: Going Beyond Multiple-Choice Tests*, workshops and conferences for teachers to promote awareness of alternative assessment options, and participation with the other nine regional labs in the development of a *Database of Alternative Assessments in Math and Science*, and a *Toolkit for Professional Developers in Alternative Assessment in Math and Science.*

In addition to these district and school level products and services, SERVE has also supported state assessment directors in the Southeast. SERVE sponsors biannual meetings of the six state assessment directors to provide a forum for them to discuss and share issues they are facing. The advent of alternative assessment and the whole discussion about moving all students to higher levels of academic performance requires that curriculum, assessment, staff development, accreditation, Special Education, and other areas work closely together in implementing changes in assessment programs. Often, there is little time and opportunity for these staff members to meet and hear from experts in the field. The biannual assessment meetings have provided an opportunity for such communication about assessment topics.

Borrowing from NCREL

ONE ADVANTAGE OF THE REGIONAL laboratory system is that labs can build upon the expertise found in other labs. This report resulted from the example set by the Regional Policy Information Center (RPIC) of the North Central Regional Educational Laboratory (NCREL). NCREL offers a series of policy papers on high stakes assessment (i.e., "the use of test results to make important decisions about the test taker"). One of these publications was entitled, "Issues and Recommendations Regarding Implementation of High School Graduation Tests" which included a report by a panel committee chaired by Dr. William A. Mehrens of Michigan State University, on the application of curricular, psychometric, educational, legal, administrative, and resource requirements for graduation tests to the Michigan context. The report was a result of a request from the state for advice on the implementation of a legislative act requiring a high school graduation test.

In the Preface to the Michigan Report it is stated:

"Certainly it is possible to develop a high school graduation test that meets curricular, psychometric, educational, legal, administrative, and resource requirements. However, as this document makes clear, the task is not easy and timelines are frequently tight. For the task to be done well, a variety of steps need to be taken soon after any legislative enactment. Immediate funding will be needed to ensure adequate human and fiscal resources. Only with appropriate funding to complete the task will

a high school test graduation requirement be of service to the citizens of a state." (pp. 16-18)

In the Executive Summary, Dr. Linda Ann Bond of NCREL provides the context for the current interest among states in implementing high school graduation tests.

"A new wave of educational reform in the 1990s has brought with it a resurgence of interest in high school graduation tests, but the types of skills that are now deemed essential to success have changed. Instead of holding students to "minimal" skills, these new mandates are intended to raise standards beyond minimal levels of achievement. Current thinking suggests that to be successful in today's technologically advanced workplace, high school graduates need skills that used to be reserved for the college-bound. Minimum competencies are not enough. Many policymakers today look to graduation tests to raise the high school graduate's skills and knowledge to the higher level expected for success in a complex, demanding society and workplace." (p. 7)

She concludes that: "Because a high school graduation test carries with it such high stakes, careful attention to the soundness of the test design process and to the legal defensibility of the test product is of critical importance." (p. 7)

The Request from Mississippi and the Response from SERVE

STATE-MANDATED TESTS REPRESENT targets set for students to achieve. As the quote from Dr. Bond suggests, these targets are moving targets. That is, expectations articulated for high school graduates in the 1970s with the first wave of high school exit tests may be different from those needed by high school graduates in the 1990s. A number of states are in the process of upgrading high school minimal competency tests developed a decade ago. Mississippi is one of those states.

Dr. Cindy Ward, the Director of Student Assessment in Mississippi, had read the NCREL report and being one year into the complexities of planning for the upgraded high school exit test felt that a panel review of their status could be very helpful as a way of ensuring that they were meeting the necessary curricular, psychometric, educational, legal, administrative, and resource requirements of a sound test development process. She approached SERVE about sponsoring such a review.

SERVE agreed to fund a Panel's review for the benefit of Mississippi's future generations who will be taking the test and for what could be learned that might help other states. Dr. Mehrens agreed to chair the panel of experts. Dr. Mehrens is a Professor of Educational Measurement and a nationally known expert in his field. He has recently been elected vice-president for the Division of Measurement and Research Methodology of the American Educational Research Association and is a past president of the National Council on Measurement in Education.

SERVE identified panel members representing a wide range of experience to make a site visit to gather information and to draft the report. Several others who were not available to make the site visits agreed to review and comment on drafts of the report.

- Two panel members, Dr. Roger Trent, the director of testing in Ohio, and Dr. Sharon Johnson-Lewis, the director of Planning, Research, and Evaluation for Detroit Public Schools, had been part of the team chaired by Dr. Mehrens which had written the report for Michigan in the NCREL document.

- The panel members had a wealth of state testing experience, including the legal challenges posed by graduation tests. State testing directors from Ohio (Dr. Roger Trent), Louisiana (Ms. Rebecca Christian), Florida (Dr. Tom Fisher), and Maryland (Dr. Robert Gabrys) participated.

- The testing expertise was balanced by curriculum expertise in the form of Mr. Lane Peeler (South Carolina Department of Education), who had been involved in the development of

curriculum frameworks in South Carolina, and Dr. Barbara Kapinus (Council of Chief State School Officers), who has worked closely with states in reviewing the implications of national content standards. In addition, Dr. Susan Barnes (Texas Education Agency) brought measurement and policy expertise through her work with the design and implementation of personnel performance assessment.

We wish to thank all of these panel participants for their willingness to take time out of their busy schedules to contribute their expertise to this project.

The Report and Its Use

AFTER THE TWO DAY SITE VISIT INcluding extensive interviews and analyses of relevant documents, Dr. Mehrens, with the assistance of the Panel, wrote an extremely informative, readable, and thorough report which applied their extensive expertise in good test development requirements to the Mississippi context, with discussions of issues and recommendations for solutions. We offer the full report in Chapter 4. Chapter 3 is the report's Executive Summary. Chapter 2 was adapted from the introduction written by the Panel and sets the Mississippi context for the report. The report concluded with the following:

- It is legally inappropriate to hold students accountable for passing an assessment that covers material that they have not been taught. This makes using a high stakes graduation assessment to drive curricular change somewhat troublesome. One can use the announcement of an upcoming assessment to drive curricular change. This, of course, requires that there be considerable time between the announcement of the assessment and its implementation.

- Multiple-choice items can measure higher-order thinking skills and procedures. Performance assessments may not offer high enough psychometric qualities to be used for high stakes assessments.

- It is unlikely that any "off-the-shelf" test would be an acceptable high school exit test for the students of Mississippi.

- Requiring any national norm-referencing component of the exit exam would complicate the task of maintaining curricular validity for the test.

- There must be close articulation among the various assessment programs. They should not work at cross purposes, and if they are serving the same purposes, perhaps less assessment is needed.

- The use of the various tests in a performance-based accreditation model requires careful thought regarding how to set the performance level and what metric to use in setting the level (e.g., average performance or percentage of students above some cut score).

The report was produced by the Panel in January of 1995. Because it was so "thorough", including 65 recommendations relative to curriculum, assessment, professional development, remediation, and accreditation, it was difficult for the assessment unit to decide how to involve others in the department in discussing the issues, especially since there was no existing interdepartmental team that dealt with assessment planning issues. Because of the number of decision points in the report, SERVE worked with the assessment department to summarize the major issues (Recommendations) in a matrix format (Table 1). This information along with the timeline (Table 2) suggested by the Panel for test implementation were proposed to the superintendent as agenda items for a meeting of involved departmental program directors.

As can be seen from reviewing Table 1, it would be extremely difficult, given the highly complex and interrelated issues involved in developing a "high stakes" exit test, for an assessment director to personally communicate to and educate others about the implications of all the decision points. The discussion guide offered a concrete way of helping the assessment staff to begin to get the

report recommendations on the table for the department.

An interdepartmental meeting of directors of Instructional Development, Accreditation, Alternative Education, Student Assessment, Title I, and Tech Prep was held in April 1995 to examine and discuss the key issues presented by the Panel Report. The SDE Interdepartmental Team met again on April 14th 1995 to discuss the issues with Dr. Mehrens, the Chair of the Panel. Thus, the report has provided a concrete means for the key state department players in the systemic reform process of upgrading expectations for student performance to come together. The importance of getting issues thoroughly discussed by such a team should not be underestimated. Such discussions early on will pay tremendous benefits in terms of the implementation of an exit test that accomplishes its objectives.

Mississippi Response to the Report

by Dr Linda Ward, Director of Student Assessment, Mississippi State Department of Education

THE DECADE OF THE 1980S, THOUGH touted as the decade of significant educational reform, established minimum standards of academic competency for students throughout the nation, and Mississippi was no exception. The Functional Literacy Examination, commonly known as the FLE, was built around specific skills identified in the Mississippi Curriculum Structure. As mandated in legislation, successful completion of this test has been required for high school graduation from Mississippi public schools since 1987. Careful attention was given to precise test development of the FLE to minimize potential legal issues which surround such a high stakes test. The test has succeeded well in fulfilling its mandate, but as desired, the passing rate on this test each year is very high. Therefore, the current FLE has been viewed by many educators and especially students as having outlived its usefulness.

At first glance, moving to a new high school exit test seemed easy enough to many. Some policymakers who were committed to the rapid improvement of instruction in Mississippi classrooms, had difficulty with the absence of visible action by both staff and a committee of practitioners charged with establishing a new exit test. After all, other components of the new Mississippi Assessment System, recommended by the Superintendent's Task Force on Accountability and Learning, including a norm-referenced test with constructed response items, were piloted in the fall of 1994.

Several committees, each charged with the responsibility of recommending implementation strategies for specific areas studied by the Superintendent's Task Force, were advancing with their work. However, curriculum revisions at various levels, along with the changes occurring from the implementation of technology into the educational system of Mississippi, promoted greater, more complex issues than those handled by other committees.

Especially challenging to the high school exit test update issue is the increasing number of assessments contained in the Mississippi Assessment System. In a state which frequently ranks near the bottom of numerous educational indicator lists, accountability takes on an even more profound nature. Results of assessments in Mississippi have demonstrated that assessment has had an impact on student performance. Yet one of the cries from educators, especially teachers, was the amount of time spent on testing with statewide assessments. Faced with many critical issues related to implementing a high school exit assessment, it seemed timely and appropriate in the fall of 1994 to seek assistance, with SERVE's help, from knowledgeable professionals outside of Mississippi.

The External Review Panel Report, which was produced from this outside review process, is comprehensive and thorough, considering only minor areas of disagreement which are inherent in the manner in which the data for the report was obtained. Of importance is the extent to which the report identified and addressed the same issues,

concerns, and problems that the High School Exit Assessment Implementation Committee has encountered. Further, the level of attention given to the interrelatedness of all assessments in the Mississippi System is significant. The report has fostered communication about important issues among internal agency offices and external parties. It has become the catalyst for continued activities related to the challenging tasks remaining in the development of a new high school exit assessment. ◆

TABLE 1

Issues in Implementing a Successful High School Exit Assessment: A Discussion Guide for Interdepartmental Planning

Decisions/Actions	*Who?*	*Funding Implications*
Curriculum/Test Specification Issues ***What standards and competencies will be assessed?*** • Readiness of curriculum frameworks to provide a basis for establishing high school exit competencies. • Tension between assessment that drives reform and ensuring the opportunity to learn. • Relationship between higher expectations and practical costs of more remediation. • Specificity of curriculum frameworks/provide sufficient direction for test specifications. • Need opportunity to learn data from students and educators prior to first pilot and again at the time of the first real administration. • Publication of competencies to be tested (notification/due process).		

Decisions/Actions	Who?	Funding Implications
• Professional development assistance for teachers to understand competencies to be assessed.		

Standards for Educational and Psychological Testing

What decisions about test development/selection need to be made?

- Customized vs. in-house.
- Who has the authority to review/accept test specifications for use in the RFP?
- Is there an instate content review team to audit the work of a contractor?
- Should a technical advisory committee offer statewide opportunities to discuss advantages and disadvantages of multiple-choice vs. performance assessment including cost estimates?

What is a reasonable timeline to ensure the opportunity to learn?

- First pilot-Fall 1996
- Second pilot-Fall 1997

Should failure rate be shared with the districts?

- First real administration-Fall 1998
- First class affected-2001 (sample attached)

Decisions/Actions	Who?	Funding Implications
What decisions about scoring/ reporting need to be made?		
• Train a standard-setting committee.		
• Will first administration or pilot data be used to get cut scores?		
• Establish a technical advisory committee.		
• Consider phased-in cut scores.		
• Establish an item sensitivity committee (bias review).		
• Discuss test vs. subtest reporting for diagnostic purposes.		
• Use a technical advisory committee to assist with equating.		
• How will personnel be trained to administer the test?		
• Consider random auditing of administration process.		
Human Resource Issues		
What recommendations should be made about fiscal needs?		
• Realistic timeline		
• Additional staff		
• Committees needed		
• Testing Policy Advisory Committee		
• Item Sensitivity Review Committee		

Decisions/Actions	*Who?*	*Funding Implications*
• Technical Advisory Committee • Three Content Review Committees • Three Cut Score Committees		

Legal Issues

What are the liability issues involved for committee members, teachers, etc.?

- What documentations/other policies need to be reviewed or put in place?
- Should a new code be written?

Accreditation Issues

What are some of the accreditation issues that need discussion?

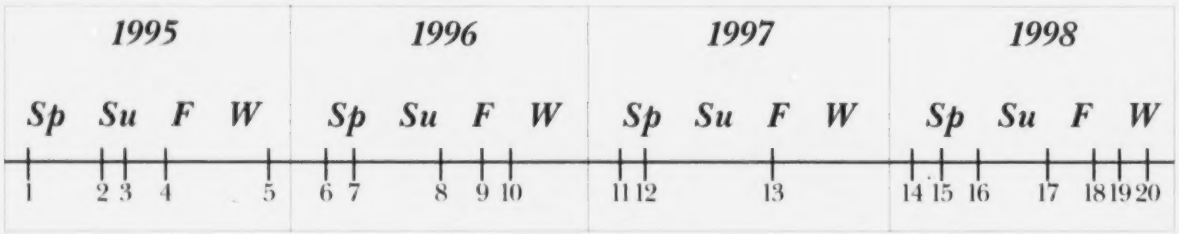
- Scaled average score and incompatible incentives
- Percent vs. average
- First try/cumulative percentage
- Alignment and weighting of standards

Coordination Issues

- Subject matter experts should study and report on the articulation of the entire testing program.
- Consider whether passing other tests are alternatives to certain MAAP tests.
- Consider whether there is too much testing.

TABLE 2

Developing the Test: A Suggested Sequencing of Tasks



Task 1: Establish appropriate advisory committees.
Department of Education Steering Committee
Testing Policy Advisory Committee
Item Sensitivity Review Committee
Technical Advisory Committee
Content Review Committees
Standard Setting Committees

Task 2: Determine what standards will be assessed.

Task 3: Disseminate information about Task 2.

Task 4: Complete test specifications for each test area.

Task 5: Hire a contractor for development of resources.

Task 6: Completion of contractor's work on resources.

Task 7: Content Committee review and revisions.

Task 8: Camera-ready copy for field testing.

Task 9: Field test items on Grade 10 students.

Task 10: Prepare and disseminate sample test items and descriptive information for teachers, students, and parents.

Task 11: Develop rules governing testing procedures.

Task 12: Analyze feedback from first field test.

Task 13: Conduct second field test.

Task 14: Revise items from the second field test.

Task 15: Select operations contractor for scoring.

Task 16: Conduct regional seminars on testing procedures.

Task 17: Complete production for first tests.

Task 18: Administer first test to Grade 10 students.

Task 19: Score and analyze results of first test.

Task 20: Design plan for releasing results to the public.

Task 21: Review and repeat steps above. This step is a continual process.

CHAPTER 2

Background Information for the Report

THE PURPOSE OF THE REPORT IS to offer readers advice on the issues that need to be considered (and resolved) and the steps that need to be taken when implementing a high school graduation test. The report also discusses some advantages and disadvantages of potential decisions. However, because Mississippi has other statewide assessments, some currently operating and others being planned, and because these other assessments interact with the high school graduation test, the External Review Panel commented to some extent on those other programs.

Mississippi Context

OBVIOUSLY, THE ADVICE GIVEN within the report is based on the Panel's understanding of the context that exists within Mississippi at the current time. Their understandings about the current context were that:

- A state code currently exists regarding the Statewide Testing Program. Relevant sections of that Code are sections 37-16-1, 37-16-3, 37-16-4, 37-16-5, 37-16-7, 37-16-9, and 37-16-11.
- In the fall of 1992, the State Superintendent of Education convened the Superintendent's Task Force on Assessment for Accountability and Learning and charged them with designing a system of assessments to serve accountability and individual assessment to meet the individual instructional needs of students. The assessments were to be designed in such a way that education would have no alternative but to change dramatically. The Superintendent was interested in implementing three major initiatives: a new assessment system, technical preparation programs, and professional development.
- The Superintendent's Task Force's report and recommendations were turned over to an Implementation Task Force. The implementation process is being carried out by three committees as follows: Norm-Referenced Assessment Implementation Committee, Workplace Competency/Employability Assessment Implementation Committee, and High School Exit Exam Implementation Committee. There exists an Overall Implementation Steering Committee that consists of the chair of the original task force, and the chair and co-chair of each implementation committee named. To-date, several committee reports have been issued.

Currently, state testing in Mississippi consists of the following:

- The ITBS and some Riverside produced and scored performance assessment exercises being

administered in grades 3-8 (for reading, language arts, and mathematics) and the Tests of Achievement and Proficiency (TAP) in grade 9 for integrated Language Arts and mathematics (pilot study this year, to be implemented in the 1995-1996 school year). These tests are to be given in the fall and used for instructional improvement and school accountability (accreditation, not broad accountability).

- The Functional Literacy Exam which is administered first in the spring to students in the 11th grade and is used both for a graduation requirement and a school accountability requirement. This is a test of basic skills in reading, written communication, and mathematics. A very high percentage (approximately 94%) of the students pass this test on the first attempt. This test is tentatively scheduled to continue at least through the 1995-1996 school year.

- The proposed Mississippi Assessment of Academic Proficiency (MAAP) (to replace the FLE). This test is to cover reading, mathematics and written communication and is to eventually be used for a graduation requirement and an accountability requirement. A set of possible instruments from different vendors was ideally (according to a "High School Exit Assessment Implementation Table") to be ready for a first pilot test to be given to 10th graders in the Spring of 1995. A test was to be selected from this set and given for standard setting purposes in the Fall of 1995. The new MAAP is not currently scheduled to "count" for high school graduation requirements or accreditation purposes until the Fall of 1996 (which would be the graduating class of 1999).

- The Subject Area Testing Program (for all students enrolled in the subject area that is tested). Currently a test exists in Algebra I, and there are plans to develop one in biology and one for U.S. History from the year 1877. The algebra test needs to be revised to fit the new curriculum structures. These tests are to be used for accountability purposes, not for making judgments about individual students.

- Occupational skills tests given to vocational completers. Currently these tests are being piloted in 15 sites and will be used for compliance with federal requirements.

- Workplace competencies to be given to students in grade 12. Currently the state is under contract with ACT to use four parts of WorkKeys (Reading for Information, Applied Mathematics, Writing, and Locating Information), and pilot administration is to begin in the Spring of 1995. The current plan is for scores on these tests to eventually be used in the accountability (accreditation) process.

The charge to the Panel primarily pertained to the new MAAP exam (to be required for graduation). However, the total context was relevant to their task. Consequently, several sections of the report will include comments related to the other portions of the statewide assessment.

Outline of the Report

THE FULL REPORT (CHAPTER 4) HAS A short section (Section I) reviewing and evaluating existing legislation and policies related to the Functional Literacy Exam (FLE). In Section II, the Panel reported on the complex issues that must be faced during the planning and implementation stages of a high school exit exam. The report calls attention to a series of issues, then recommends solutions to some of them. In writing this section, the panel tried to reference current procedures for the FLE because many of the procedures would be the same for the proposed MAAP, or indeed any exit examination. Nevertheless, the panel acknowledged that they were probably not fully aware of all current procedures and, therefore, may in parts be offering advice that the Mississippi Department of Education has already implemented. They hoped that the Department personnel would feel complimented rather than offended by suggestions about procedures and policies already in place.

Section III provides an overview of some of the important steps to be considered in developing and implementing a high school graduation test and suggests when these steps need to be taken. Again, parts of this section may be offering suggestions that are already being implemented with the FLE and can simply be adapted for the MAAP.

Section IV relates to issues that should be considered when using test results for accreditation purposes. Although the Panel did not receive a specific request for this section, they felt that discussions about the use of test results were critical.

Obviously, discussions in this report such as the specific procedures to follow in building and implementing an exit test, their timing, and the resolution of the issues often overlap. It needs to be stressed that no procedure will produce a perfect assessment instrument or process. Perfect assessment procedures simply do not exist. However, a test should be as good as it can be given the constraints. Whether any given test or process is legally defensible is ultimately a decision for the courts. If followed, standards established by the measurement profession make a test more defensible. But, no set of standards should be used as a checklist.

This report cannot (and is not intended to) replace the advice that a state department of education will need from an ongoing technical advisory committee. The advice from such a committee is essential to the development of a technically and educationally sound program.

Definition of Some Terms

THERE ARE SOME MEASUREMENT terms which, while they have fairly standard definitions among measurement experts, are not always used by other educators and lay people with the same meanings. To facilitate communication, the panel provided the following definitions of some commonly used terms.

Norm-referenced

When an assessment is norm-referenced, the scores made by individual students are compared to the scores of some identified norm group or groups. The norm group may be local, state, national, or (theoretically) global. A requirement of norm referencing is that the assessment given to all those who wish to reference scores against the norm group be given under the same standardized administrative conditions and scoring procedures as the original norm group. For example, if students in the norm group took the assessment under timed conditions, the students whose scores were to be compared to the norm group would have to take the test under the same time constraints. Norm-referenced scores are not dependent upon any type of item format. Performance assessments and multiple-choice tests can both be norm-referenced. Norm referencing scores does not prohibit setting standards or employing criterion-referenced test interpretation.

Criterion-referenced

In criterion-referencing, one references the scores by comparing them to a standard or set of standards. In a high school exit exam there would typically be one standard and students whose scores were at or above the standard would pass and those whose scores were below the standard would fail. As with norm referencing, criterion-referencing does not depend upon any type of item format. Scores from both multiple-choice and performance assessments can be criterion-referenced. Again, to be fair, all students should take the test under the same administrative and scoring procedures. However, if there is no external norm group, one need not be concerned about standardizing the conditions to those under which the external group took the test.

Multiple-choice Item

Obviously this type of item has a set of options, and the test takers choose (typically) one of the options as the best answer. These items can be machine scored very quickly. Contrary to some rhetoric, multiple-choice items can assess higher-order thinking skills and indeed can require problem-solving skills to obtain the correct answer.

Performance-based Assessment

While this term can be used in a variety of ways, we will use it to mean all assessments that require some sort of constructed response which needs to be scored. Performance-based assessments can, but do not necessarily, require higher-order thinking skills or problem-solving abilities. As mentioned above, to be fair, all students should take these performance exercises under the same administrative and scoring procedures. ◆

CHAPTER 3

Executive Summary of the Full Report

Issues and Recommendations Regarding the Mississippi Assessment of Academic Proficiency (MAAP) and Other Components of their Assessment for Accountability and Learning Plans

A report prepared for the SouthEastern Regional Vision for Education (SERVE), the Mississippi High School Exit Assessment Implementation Committee, the Mississippi State Board of Education, and the Mississippi Department of Education by an External Review Panel.

Chair:

Dr. William A. Mehrens, Michigan State University

Panel Members:

Dr. Susan Barnes, Texas Education Agency
Ms. Rebecca Christian, Louisiana Department of Education
Dr. Sharon Johnson-Lewis, Detroit Public Schools
Mr. Lane Peeler, South Carolina Department of Education
Dr. Roger Trent, Ohio Department of Education

Other Reviewers:

Dr. Thomas Fisher, Florida Department of Education
Dr. Robert Gabrys, Maryland Department of Education
Dr. Barbara Kapinus, Council of Chief State School Officers

THIS REPORT OFFERS READERS advice on the issues that need to be considered (and resolved) and the steps that need to be taken when implementing a high school graduation test. Following a general introduction and a review of legislation and policies related to the FLE, issues are discussed under the following headings:

- Curriculum/Test Specification Issues,
- Additional Curriculum and Instructional Considerations,
- Psychometric Issues,
- Education Issues,
- Legal Issues,
- Policy/Administrative Issues, and
- Human and Financial Resource Issues.

Following those discussions, we include a short section on the sequencing of tasks and a section on using test scores for accreditation purposes.

The report contains 65 different recommendations surrounded by extensive discussions. However, the report cannot (and is not intended to) replace the advice that the State Department of Education will need from an ongoing technical advisory committee. Some of the more important of the issues and recommendations are discussed below in this executive summary. However, we urge all readers to study the total report.

Curriculum/Test Specification Issues

AS RECOMMENDED BY THE HIGH School Exit Assessment Implementation Committee, the initial assessment areas of the new MAAP should be limited to reading, mathematics and written communication. There needs to be much more evidence of curricular validity (opportunity to learn) prior to the implementation of the new MAAP and the state has an obligation to provide professional development to local teachers regarding how best to ensure that the new competencies are adequately taught.

We support the development of additional curricular structures. We believe the term curricular structures is preferable to frameworks.

Psychometric Issues

VALIDITY REFERS TO THE DEGREE TO which evidence supports the inferences that are made from assessment scores. Department of Education employees need to be cautioned against making any unsubstantiated statements about what the assessment measures or what inferences can be made from the assessment scores. For example, a statement such as the assessment would ensure that if students passed they "would be able to be successful in the real world," implies evidence of predictive validity. If no predictive validity exists, such inferences cannot legitimately be drawn.

Exercise development is very important. If the developed exercises are faulty, the assessment will be inadequate. The assessment should be specifically constructed for the graduation requirement. It is very unlikely that any "off the shelf" existing assessment package would be adequate. A Request for Proposal (RFP) should be issued to develop this assessment package. The RFP should demand that the contractor design sufficient safeguards into the assessment development to ensure adequate content validity. Both department employees and an in-state content review team should be involved in reviewing various processes and products throughout the development stage.

Performance assessments should be used with care and recognition and consideration should be given to the fact that such assessments are frequently not as psychometrically sound nor as cost effective as the more traditional multiple-choice assessments. It is true that performance assessments can assess some competencies that cannot be assessed with

multiple-choice items, and they should be used for such competencies. Nevertheless, multiple-choice items are the most efficient and effective format to assess many competencies. There need to be compelling reasons for using performance measures to assess those competencies that are amenable to assessment with more traditional approaches.

We applaud the recommendation of the High School Exit Assessment Implementation Committee that there be two years of pilot work on the new MAAP. However, we think their proposed time-line is too optimistic. We advise that Mississippi move a bit more slowly than originally planned, proceed with appropriate thoroughness, and document every step of the design and implementation process.

Many specific recommendations are made regarding technical issues such as scoring, standard setting, item sensitivity reviews and item bias studies, reliability, scaling/reporting, the number of forms required, equating, and standards of test administration. Some of these are quite technical in nature and will not be covered in this executive summary.

Education Issues

THE NEW MAAP NEEDS TO BE ARTICULated with the other tests in Mississippi. Thought needs to be given as to whether the various tests and their specific uses within the state complement each other or result in competing goals. Specific procedures regarding retesting should be planned and adopted by the Board. A proposal that addresses questions regarding the remediation efforts and the respective responsibilities of the state, the district, and the student needs to be developed.

Legal Issues

LIABILITY ISSUES MUST BE CONSIDERED. Necessary statutes with respect to liability should be obtained. All committees and staff should be informed regarding their potential liability.

Students and their parents need to be given sufficient notice regarding the new graduation requirement. The new MAAP should not be implemented until it can be demonstrated that students have had an opportunity to learn the competencies to be assessed.

All procedures, security provisions of the assessment, and issues concerning accommodations must be documented.

Policy/ Administrative Issues

A PLETHORA OF ISSUES NEED TO BE resolved including administrative rules, frequency of administra ion, etc.

Human and Financial Resource Issues

WE CAUTION AGAINST PROCEEDING without sufficient staff and resources. There probably needs to be additional staff in both the student assessment and the curriculum/instructional units of the department. Advisory committees need to be established. These include a testing policy advisory committee, an item sensitivity review committee, a technical advisory committee, a content review committee for each content area of the assessment, and a committee to recommend the cut score. The number of contractors should probably be limited to two. Sufficient financial resources are needed to do the high quality job required to build an educationally sound and legally defensible assessment. Information from other states should be obtained to assist in determining the amount of resources needed.

Using Test Scores for Accreditation Purposes

IN THE ACCREDITATION SYSTEM, THE "success of the school system" could and perhaps should be defined in terms of the number of students who demonstrate the desired level of performance rather than in terms of average scores. It may be preferable to use the cumulative proportion who have passed the MAAP at the end of some given grade (e.g., grade 11), rather than the initial pass rate. The MDE should study carefully the alignment and weighting of all performance standards used across the elementary, middle, and high school grades.

Conclusions

IT IS POSSIBLE TO DEVELOP A WELL-designed high school graduation test that meets curriculum, psychomet ric, educational, legal, administrative, and resource requirements. However, the task is not easy. For it to be done well, a variety of steps need to be completed. For those steps to be completed, adequate funding must be made available.

While all the recommendations are not covered in this executive summary, we point out below some of the most pertinent aspects that have been considered in the report.

- It is legally inappropriate to hold students accountable for passing an assessment that covers material that they have not been taught. This makes using a high stakes graduation assessment to drive curricular change somewhat troublesome. One can use the announcement of an upcoming assessment to drive curricular change. This, of course, requires that there be considerable time between the announcement of the assessment and its implementation.

- Multiple-choice items can measure higher-order thinking skills and procedures. Performance assessments may not offer high enough psychometric qualities to be used for high stakes assessments.

- It is unlikely that any "off-the-shelf" test would be an acceptable high school exit test for the students of Mississippi.

- Requiring any national norm-referencing component of the exit exam would complicate the task of maintaining curricular validity for the test.

- There must be close articulation among the various assessment programs. They should not work at cross purposes, and if they are serving the same purposes, perhaps less assessment is needed.

- The use of the various tests in a performance-based accreditation model requires careful thought regarding how to set the performance level and what metric to use in setting the level (e.g., average performance or percentage of students above some cut score). ◆

CHAPTER 4

The Full Report

Issues and Recommendations Regarding the Mississippi Assessment of Academic Proficiency (MAAP) and Other Components of their Assessment for Accountability and Learning Plans

January, 1995

A report prepared by an External Review Panel for the SouthEastern Regional Vision for Education, the Mississippi High School Exit Assessment Implementation Committee, the Mississippi State Board of Education, and the Mississippi Department of Education.

Written by
William A. Mehrens
Michigan State University, Chair of the External Review Panel with great assistance and final agreement by the other members of the External Review Panel

External Review Panel Members

Dr. William A. Mehrens, Panel Chair
Professor of Educational Measurement
462 Erickson Hall
Michigan State University
East Lansing, MI 48824-1034

Dr. Sharon Johnson-Lewis,
Assistant Superintendent
Department of Research, Development and Coordination
Detroit Public Schools
5057 Woodward Ave. Suite 776
Detroit, MI 48202

Dr. Roger Trent, Director
Division of Educational Services
Ohio Department of Education
65 South Front Street, Room 811
Columbus, OH 43266-0308

Mr. Lane Peeler
Division of Curriculum and Instruction
South Carolina Department of Education
1429 Senate Street
Columbia, SC 29201

Dr. Susan Barnes
Division of Policy
Texas Education Agency
1701 N. Congress, Room 3-104
Austin, TX 78701

Ms. Rebecca Christian, Director
Bureau of Pupil Accountability
Louisiana Department of Education
1366 Dahlia
Baton Rouge, LA 70808

Other Reviewers of the Document

Dr. Thomas Fisher,
Administrator
Student Assessment Services
Florida Department of Education
325 West Gaines Street
Room 701
Tallahassee, FL 32312-0400

Dr. Robert Gabrys
Planning, Results, and Information Management
Maryland State Department of Education
200 W. Baltimore St.
Baltimore, MD 21201

Dr. Barbara Kapinus
State Education Assessment Center
Council of Chief State School Officers
One Massachusetts Avenue, N.W.
Suite 700
Washington, DC 20001-1431

TABLE OF CONTENTS

Foreword 22

Section I: Review of Legislation and Policies Related to the FLE 23

Section II: Issues and Recommendations Regarding the Proposed MAAP 23

Curriculum/Test Specification Issues 24
- *Specify Subject Matters* 24
- *Specify Content within Subjects* 24

Additional Curriculum and Instructional Considerations 26
- *Mississippi Curriculum Structures* 26
- *Mathematics Curriculum Structure* 27
- *The Functional Literacy Examination (FLE)* 28

Psychometric Issues 28
- *Validity* 28
- *Exercise Development* 29
- *Mix of Item Formats* 30
- *Pilot (field) Testing* 30
- *Scoring* 32
- *Standard Setting* 32
- *Item Sensitivity Reviews and Empirical Bias Studies* 34
- *Reliability* 35
- *Scaling/Reporting* 35
- *Number of Forms* 37
- *Equating* 37
- *Standardization of Test Administration* 37

Education Issues 38
- *Articulation with Other Tests* 38
- *Retesting* 39
- *Remediation* 39

Legal Issues 40
- *Liability issues* 40
- *Notification* 40
- *Timing* 40
- *Opportunity to Learn* 41
- *Documentation* 41
- *Security Provisions* 41
- *Accommodations* 42

Policy/Administrative Issues 42

Human and Financial Resource Issues 43
- *Staffing Needs* 43
- *Advisory Committees* 43
- *Contractors* 44
- *Financial Resources* 44

Section III: Sequencing of Tasks 44
- *Sample Tasks and Completion Dates* 45

Section IV: Using Test Scores for Accreditation Purposes 47

Section V: Conclusions 49

References 50

End Notes 50

FOREWORD

THE EXTERNAL REVIEW PANEL on the Mississippi Assessment of Academic Proficiency (MAAP) was convened to advise the Mississippi High School Exit Assessment Implementation Committee (HSEAIC), the State Board of Education (MSBE), and the Mississippi Department of Education (MDE) on important issues surrounding the proposed high school proficiency examination and other components of their state assessment programs. The panel members are national experts who have first-hand knowledge and experience with large-scale testing programs; they brought to the task a wealth of information and wisdom on the challenging issues that Mississippi educators will face as they develop and implement different aspects of their proposed new assessment programs.

Our specific charges were as follows:

- Review and evaluate legislation and policies specifically related to the current high school exit exam in Mississippi, the Functional Literacy Exam (FLE).
- Review and evaluate the process and procedures for designing and identifying and/or determining the new high school exit assessment, the Mississippi Assessment of Academic Proficiency (MAAP).
- Review and evaluate curricular and instructional documents related to statewide assessments at the secondary level, especially those more closely related to the FLE, and including proposed academic standards and competencies issued to districts for review for MAAP.
- Conduct a two-day site visit in Jackson, Mississippi, consistent with objectives approved by the High School Exit Assessment Implementation Committee; and
- Complete a post-visit report on the status of the project, to include recommendations to the High School Assessment Implementation Committee and the State Board of Education.

All six members of the External Review Panel had been mailed a package of materials on the background and implementation plans for the proposed Mississippi state assessment programs. Three members of the External Review Panel met in Jackson, Mississippi on November 30, December 1, and December 2, 1994 (one additional member could only attend on December 2). On December 1 and 2, those members of the External Review Panel present had the opportunity to interact with individuals from the following groups: Mississippi Department of Education Administrative Staff, MDE Assessment Staff, the Superintendent's Task Force on Accountability and Learning, Curriculum specialists, and the High School Exit Assessment Implementation Committee. In addition, we met with the Director of the Education Forum of Mississippi. During those two days we also received other printed materials related to the state assessment programs.

The other two members of the panel planned to attend the meetings in Jackson, but emergencies kept them from doing so. Nevertheless they read all the materials sent them before and after the meetings and have read, reacted to, and agree with this final report.

Three other national experts have reviewed this panel's report for SERVE and their comments have been considered and basically followed in the final draft of this report. They are obviously not responsible for any errors in the report, and this report should not be considered as having been endorsed by them (although it is our belief that they basically agree with this report).

It is important to note that portions of the outline and indeed much of the general content of this report is patterned after *Issues and Recommendations Regarding Implementation of High School Graduation Tests*, written by William A. Mehrens for the North Central Regional Educational Laboratory (Mehrens, W.A., 1993). That report, in turn, was patterned after a report written by an expert panel, and chaired by William A. Mehrens, for the Michigan Department of Education. We appreciate the consent of the NCREL to use that material.

SECTION I:

Review of Legislation and Policies Related to the FLE

WE HAVE READ SEVERAL PORTIONS of Section 37 from a Code related to Statewide Testing Programs and some Department reports on the FLE (June, 1993) and the total Mississippi Statewide Testing Program (Summary Report for 1994) that inform us (although probably only partially) regarding relevant legislation and policies related to the FLE. With respect to the Code, various portions of Section 37 seem appropriate legislation for the FLE and indeed for the proposed MAAP. A February 10, 1987 letter from [then] state superintendent Boyd to the Honorable Jack Gordon correctly pointed out that, . . . "it is cumbersome and untimely to seek legislative amendments for what are routine educational decisions." It appears that the legislators agreed with this position and the Code is reasonably general. Because the plan for the MAAP is to assess the same three areas (reading, writing, and mathematics), the code may be appropriate for the MAAP although written for the FLE. However, in various sections of the Code, the terms "basic skills" and "functional literacy examination" (lower case, not as a name of an exam) are used. It would probably be preferable if a new code were written with language more in keeping with the proposed MAAP. Additionally, we recommend that someone from the Attorney General's Office review sections 37-16-9 and 37-16-11 to see if they are worded appropriately given new federal legislation (such as the Americans with Disabilities Act [ADA]) and any new state legislation or legal precedents.

In summary, we recommend that the code be reviewed to ensure that it is appropriate, based on the proposed MAAP and the new federal legislation. Furthermore, until more details are decided with respect to the MAAP, it is difficult to determine just which policies or procedures should be decided by the Department (and confirmed by the Board) and which may need to be in legislative code. Thus, our first recommendation.[1]

Recommendation 1:
Engage the services of someone from the attorney general's office regarding whether a new (revised) code needs to be written. The relationships with this individual should be ongoing so, as more specific decisions are made about the MAAP, that individual can have input into the decision of whether board policies or Legislation is preferable.

With respect to specific policies regarding the FLE that have been formulated by the State Department of Education, those of which we are aware seem well conceptualized and have served the state well with respect to the FLE. In the section to follow, we will be making some specific suggestions with respect to policies that should be adopted for the MAAP. To the extent existing policies already cover our suggestions, they can be ignored.

SECTION II:

Issues and Recommendations Regarding the Proposed MAAP

MANY ISSUES MUST BE CONSIDERED when implementing a high school graduation test. This section will address several of the more important ones, including curriculum/test specification, psychometric, educational, legal, policy/administrative, and human/financial resource issues. Many of the issues are connected and the resolution of one may affect the others.

In preparing this report, we were mindful of legal and professional guidelines that must be considered when designing and implementing a required high school graduation test. Professional standards for tests are articulated in *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985). Many of the legal considerations have been addressed in the case of *Debra P. v. Turlington* (1983,

1984), a broad-based challenge to Florida's high school graduation test requirement.

Curriculum/Test Specification Issues

OBVIOUSLY, ONE MUST DECIDE WHAT to test before beginning to construct the test. But the task is not a simple one. General decisions need to be made regarding what subject matters to test, but more specific decisions need to be made also, including what subareas to test in those subject matters and how many questions should come from each of the subareas. These decisions are important for educational, psychometric, and legal reasons. This section discusses and offers recommendations on some of the more important issues.

Specify Subject Matters

The high school exit assessment implementation committee has recommended that "testing will be only in the areas of reading, math and written communication because these subjects measure some important basic high school competencies." We agree with this recommendation. To add additional areas would increase the costs and make time-lines more difficult to meet. However, the state may wish to add additional subject matter areas in the future.[2]

Recommendation 2:
The state board should abide by the recommendation of the implementation committee to limit the exit assessments to the areas of reading, mathematics, and written communication. Additional areas may be added at a later date.

Specify Content within Subjects

After deciding which subject areas to assess,one must decide how those subject matters are to be defined and which particular subparts to assess. In keeping with the terminology used in Mississippi, one must determine what standards and competencies should be assessed. Obviously, high school graduation tests should not sample a state's total curriculum for measurement, philosophical, and legal reasons.

A particularly troublesome problem in Mississippi is that the Curriculum Structures are in various stages of revision. For example, the Mississippi Mathematics Curriculum Structure has a 1995 date on the outside cover and an October, 1994 date on the inside cover page. The official mandated reading and written communication skills curricula were published in 1986, and a revision process is scheduled to begin in early 1995. As we understand the current plans, the English/Language Arts and Reading will be meshed into one curriculum.[3] The revision is estimated to take about 18 months. Current plans (hopes) are that the curriculum structures will be completed in time for training in the summer of 1996, the curriculum structures will be piloted in 1996-1997 and will be implemented first during the 1997-1998 school year.

Thus, schools are currently required to teach the 1986 curricula in reading and written communication skills, and they have just been mandated to teach the revised mathematics curriculum.

A second problem has to do with the articulation of both the content and the timing of the old FLE and the proposed MAAP. (We will make subsequent recommendations following this articulation when we discuss the timing concerns.) The FLE, still required for high school graduation and still being used for accountability purposes, covers content that the schools should feel obliged to teach. The FLE covers subparts of the 1986 curricula. Implementing a new MAAP that covers different (or perhaps just additional) competencies leaves schools in a quandary regarding which of the competencies should have higher priority in their curriculum.

The process used thus far in Mississippi to determine the specific standards and competencies that are to be the basis of the MAAP is considerably less than exemplary and far from complete. In fact, one member of the implementation committee has suggested that the process has been conducted in somewhat of a backwards fashion.

As we understand the process, the implementation committee divided into three groups and, with a couple hours of work, produced a set of standards and competencies in Mathematics. These basically came from the brand new Mathematics Curriculum Structure. For the other two areas they were produced based on the committee members' knowledge of national movements in these two curricular areas. After these statements had been produced, they were put into a survey (the Academic Content Standards & Competencies Questionnaire) and mailed to school districts. For Reading and Language/Written Communication the districts were asked to respond yes or no to three questions for each Standard and Competency: Currently being taught, Should be part of curriculum, and Should be tested. Because there exists a revised curriculum in Mathematics and the standards and competencies came from that, the districts were only asked whether they should be tested. (We assume that there is other documentation regarding the responses of schools to the first two questions.) The results of the survey are contained in a document entitled "Analysis of Results on the Academic Standards and Competencies Questionnaire Administered During September 1994" dated October 21, 1994. In general the results suggested that "all of the standards and competencies in reading are currently being taught, should be part of the statewide curriculum, and should be measured on the exit examination." (Percents for Yes responses ranged from 77% to 100% across all three questions for all standards and competencies.) For Language/Written Communication, "most of the competencies are being taught and should be part of the curriculum. However, four standards and one competency received low percentages of 'yes' responses on the question whether they should be measured on the exit exam." For Mathematics, "the responses indicated that the reviewers felt that one standard and three competencies in the current curriculum should not be measured on the exit examination."

We commend the gathering of the data on the competencies and standards. As we understand it, the standards and competencies for all three areas will be revised based on the responses to the questions and the open-ended responses.

In the deliberations regarding the revisions of the standards and competencies, the HSEAIC needs to balance two competing interests. In his original charge to the task force, the Superintendent wanted the assessment program "to be designed in such a way that education will have no alternative but to change dramatically." This charge, coupled with the moral and legal needs for holding students accountable for learning only if they have had the opportunity to learn presents a clear tension between two existing forces. This tension is not unique to Mississippi. Around the country those interested in reforming education have suggested (correctly) that assessment can serve as a powerful catalyst for educational change. If schools and/or students are held accountable for certain standards and competencies, then those will be taught. The tension arises because it is morally reprehensible and legally impermissible not to grant a high school diploma (a property right) to students because they have not learned material that has not been a part of their curriculum. In the *Debra P. v. Turlington* precedent, it was held that a student cannot be denied a high school diploma *unless it has been adequately demonstrated that the student has had an opportunity to learn the material on the test.* This legal precedent has been incorporated into the professional *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985):

> When a test is used to make decisions about student promotion or graduation, there should be evidence that the test covers only the specific or generalized knowledge, skills, and abilities that students have had the opportunity to learn (p. 41-42).

Thus, tension exists between wanting to use a high stakes assessment (for students) to reform education and the desire to be fair to the students and to have a legally permissible assessment. The state must consider carefully the trade off between (a) using assessment of desirable but currently not-taught material as a catalyst for curricular change and (b) restricting exit assessment content to material that has been taught. While we recognize the tension, we lean toward the fair and legal side rather than the catalyst for curriculum reform side of the debate.[4] Thus, we offer the following recommendations.

Recommendation 3:
There should be another survey of the local

districts to determine the opportunity to learn the standards and competencies prior to the first pilot administration of the test, and again at the time of the first real administration. We strongly recommend that both students and educators be surveyed. If the evidence from the opportunity to learn surveys suggests that the material to be on the tests has not been adequately covered in the curriculum, we suggest the exit requirements for the assessments be postponed.

Recommendation 4:
Once the specific standards and competencies are determined, this information should be widely publicized in the local school districts. This information should be disseminated in enough detail to make students, parents, and educators aware of the knowledge and skills to be tested without providing so much detail that the students can answer the questions without understanding the curriculum.

Recommendation 5:
If the assessment is to include any material not currently mandated by the state or taught in the schools, there should be a state board administrative rule or statute which specifies that the local districts must teach this material.

Recommendation 6:
Once the standards and competencies are determined, the state must provide assistance in the professional development to local teachers if there is a need.[5]

Additional Curriculum and Instructional Considerations

WE WOULD LIKE TO RAISE SEVERAL additional issues concerning Mississippi curriculum structures in general and the mathematics curriculum structure in particular. The mathematics curriculum structure is singled out because it is complete and because mathematics will be one of the tested areas at lower levels and on the new MAAP.

Mississippi Curriculum Structures

It is the understanding of the External Review Panel that all of the core subject areas eventually will have curriculum structures. The mathematics curriculum structure and its accompanying process guide as well as the curriculum structure for social studies were given to the panel for review. The science curriculum structure will soon be considered for state adoption by the State Board of Education according to the science specialist. The English/Language Arts/Reading curriculum structure will be interdisciplinary in nature and will be developed over the next 18 months. In all cases it was reported that the curriculum structures attempted or will attempt to embody the current thinking of national professional organizations and documents. Obviously this is desirable in that national publications and conferences will address various issues that are pertinent to the Mississippi situation.

In several instances the term "framework" was used by various persons interviewed as a synonym for curriculum structure in a particular subject area. The term "curriculum structure" appeared on both the mathematics and the social studies documents reviewed. Based on the composition and organization of other state and national frameworks, the Mississippi curriculum structures are more akin to curriculum guides. Curriculum frameworks are broad in scope and do not provide teachers with specific objectives for use in their classrooms. Frameworks communicate the spirit, not the specifics, of the mathematics curriculum. In addition, frameworks address other issues such as professional development, instructional materials adoption procedures, essential support systems, etc. The two Mississippi curriculum structures reviewed do provide information to the objective level and define the curriculum in fairly specific ways. Hence the term curriculum structure is the more appropriate term to use when referring to the documents that have been or are being developed to guide instruction in Mississippi schools.

Recommendation 7:
The current documents used to define the Mississippi curriculum in appropriate subject areas should not be referred to as frameworks because they are more specific than frameworks in the area of content but do not fully address other areas that frameworks generally do. Curriculum structure is the term that appears to be the choice of Mississippi educators for the documents described above.[6]

Mathematics Curriculum Structure

Because the mathematics curriculum structure will be used to define the mathematics content that will be tested on the MAAP as well as other measures of achievement, further analysis will be made of that document and its companion document, the mathematics process guide. Both of these have recently been completed and some teacher orientation sessions held during the fall of 1994.

Six content strands have been identified in the *Mississippi Mathematics Curriculum Structure* as strands that will be taught at every grade level. These are as follows: number sense/numeration/operations, patterns/relations/functions, algebra, measurement, geometry, and statistics/probability. Objectives are identified for each strand for each grade level K-8 and for specific courses in the upper grades' curriculum. Prealgebra can be taught as early as grade 7, and algebra can be taught as early as grade 8. The objectives for each strand seem to fit adequately into the grades K-8 curriculum since each strand will likely be addressed at each of these grades. Some of the objectives in some of the upper level courses seem "forced." While it is true that these six strands will be addressed in the secondary curriculum, it may not be true that each of the strands will appropriately fit into each course of the secondary curriculum. For example, the statistics/probability strand seems to be one that does not fit well in each of the courses. Specifically, for A.P. Calculus, the objectives under the statistics/probability strand (p. 77) are as follows: solve related rate problems, solve optimization problems, and use integration to solve real life problems. The connection of the objectives with the strand is not readily apparent. Hence, it may not be true that every strand will conveniently fit into every course.

Recommendation 8:
We suggest that the mathematics curriculum structure be reexamined in light of the above issues. Each strand may not fit well into every course at the secondary level and need not be forced to do so. Hence, as tests are constructed for various purposes, this issue should be recognized.

The seven process strands identified in *The Mississippi Mathematics Process Guide* are as follows: problem solving, communicating, reasoning, connecting, estimating, using technology, and assessing. If the content strands identified earlier were written horizontally across the top of a grid and the process strands were written down the left side of the same grid, each cell of the grid would represent the interaction of a content strand and a process strand. In general, this is the kind of information that the process guide provides for each of grades K-8. Each cell would contain an activity that indicates how a content strand and a process strand might interact in a desirable fashion. For Prealgebra, Algebra I, Geometry, Algebra II, and Trigonometry, activities are described which generally reflect combinations of several of the process strands to some topics in those courses.

It appears that the content selected (through the activities) for interaction with the process strands is not always that delineated in the objectives of the mathematics curriculum structure. In general the link between the process strands of the process guide and the content objectives of the curriculum structure are not always explicit. We believe that this is the crucial link about which teachers are seeking help. How well teachers understand this relationship will determine how the content is presented in classrooms.

Recommendation 9:
It would be desirable to have a closer correlation between the objectives of the curriculum structure and one or more of the process guide. Teachers could then better understand the "what" and the "how" of their curriculum. Because both documents have already been published, some emphasis should be given to this connection between process and content strands in staff development sessions at all levels. The

BEST COPY AVAILABLE

degree to which all of the curicular and assessment emphases correlate to one another at all levels should be considered.

Of all the tests, only the MAAP (eventually) can deny a student a diploma. However, all of the other assessments should be providing "help" for students to do well on the MAAP (see recommendation 62).

The Functional Literacy Examination (FLE)

Obviously since the composite passing rate of the current FLE is about 94%, the material being tested has been taught by teachers and is being comprehended by almost all students. The new MAAP will likely assess more advanced content when it is ready. However, in the interim, teachers are being asked to adjust their curriculum and their teaching to correspond to the mathematics curriculum structure and the mathematics process guide, but passing the FLE will remain a graduation requirement. Some attention should be given to how the current FLE content will be incorporated into the new initiatives so that students will continue to perform well, even while addressing the new curriculum structures, until the new MAAP is ready.

Psychometric Issues

ALL PARTICIPANTS IN THE TEST CONstruction, administration, scoring, and reporting process should be aware of the *Standards for Educational and Psychological Testing* mentioned earlier. This section is divided into subsections on validity, item development, mix of item formats, field testing, scoring, standard setting, item sensitivity reviews and bias studies, reliability, scaling/reporting, number of forms, equating, and standardization of administrations.

Validity

> Validity is the most important consideration in test evaluation...[and] refers to the degree to which that evidence supports the inferences that are made from the scores (AERA, APA, NCME, 1985, p. 9).

Although validity is a unitary concept, evidence of validity may be accumulated in many ways. Traditionally, such evidence has been categorized as content, criterion-related, and construct validity evidence. Different inferences that may be drawn from a test score demand different types of validity evidence. It is important not to make insupportable inferences from the scores. The test name itself may lead to an insupportable inference. For example, calling a test a "Functional Literacy Examination" as has been done for the previous exit examination would support the inference that a person who failed the test was illiterate. Thus, the name should be chosen with care. *We support the suggestion made by the HSEAIC to call the test the Mississippi Assessment of Academic Proficiency.*

In addition to what a test is called, it is important that public officials do not suggest in their writings or speaking that inferences can be drawn from the assessment which are not supportable. For example, in our meetings on December 1 and 2 we heard one person suggest that the assessment would ensure that if students passed they "would be able to be successful in the real world." Such a statement implies that there is some evidence of predictive validity. If no predictive validity exists, such inferences can not legitimately be drawn. Thus, the following recommendations.

Recommendation 10:
Every effort should be made to caution Department of Education employees, the State Board of Education, and other spokespersons against making any unsubstantiated staements about what the assessment measures or what inferences can be made from the assessment scores. An official statement shoul be made regarding the assessment and the inferences that can be drawn from the scores. There should be either good logical reasons or empirical evidence for the inferences that are to be drawn.

Recommendation 11:
Professional development activities related to the new test should include discussions about valid statements that may be made about the test results.

Exercise (test item) Development

If the developed exercises are faulty, the assessment is inadequate. One of the most important aspects of a good assessment is that it, indeed, measures the standards and competencies (hereafter just called content) that have been listed in the publications describing the assessment content. One of the major Standards to be considered is as follows:

> When a test is to be used to certify the successful completion of a given level of education...both the test domain and the instructional domain at the given level of education should be described in sufficient detail, without compromising test security, so that the agreement between the test domain and the content domain can be evaluated (AERA, APA, NCME, 1985, p. 52).

This evaluation should not be left for the test's critics to make after the test has been given. This evaluation needs to be made at the time an assessment is chosen or developed. Ensuring the test/curriculum match and communicating the test domain to others is likely to be more difficult if the curriculum structures stress quite broad, general competencies. For example, one proposed curriculum structure (science) will list competencies as well as sample objectives for the local districts to cover. A problem is that the test will probably assess at the objective level, and some districts could be instructing to objectives that match the broad competencies on the curriculum structures but not the specific objectives assessed on the test.

In ensuring a match between what the assessment measures and the publicized content, it is unlikely that any "off the shelf" existing assessment package would be adequate. Most likely, an assessment package will need to be specifically built to match the Mississippi standards and competencies.

> ***Recommendation 12:***
> ***Plan on constructing an assessment to be used specifically for the graduation requirement. Be very skeptical of any contractor who suggests an off the shelf test will adequately meet the requirements of a Mississippi High School exit examination.***

In developing the assessment, there must be several steps taken to ensure an adequate match between content and test specifications.

> ***Recommendation 13:***
> ***Demand that the contractor design sufficient safeguards to ensure that the assessment adequately samples the defined content.***

Items can be faulty for a variety of reasons. If the original items are faulty, either because they do not match the defined content or for other reasons, it is difficult to "fix" the test at the field test stage of development. Any item substantially revised following a field test should be subjected to another field test. Thus, it is important to have well-trained item writers.

> ***Recommendation 14:***
> ***Any request for proposal (RFP) for item/test development must be written to elicit sufficient information from the prospective contractors so that the bid will not be awarded to an incompetent contractor. The department will need to audit closely the work of the contractor to ensure adequate item development, tryouts, revisions, etc. It is critically important to have an instate content area review team composed of teachers, curriculum supervisors and university curriculum specialists determine the quality of the item specifications and the items and recommend appropriate revision to the contractor.[7]***

Because the FLE has been given in the same three subject matter areas called for in MAAP, it may be possible that some of the items used in that exam can be used (or revised) for the MAAP. Given the expense of item development, it would be wasteful to discard items if they match the new standards and competencies.

> ***Recommendation 15:***
> ***Both the department and an instate content review team should review the items currently being used on the FLE to determine whether any of them match the new standards and competencies. If they are of sufficiently good quality, consider using those items on the MAAP.[8]***

Mix of Item Formats

An issue that should be considered early is the mix of item formats. Some critics of multiple-choice items suggest incorrectly that such a format cannot tap into higher-order thinking skills. The rhetoric of such critics suggests that multiple-choice items can only measure basic, isolated bits of recall. Such is clearly not the case and we hope that those designing the assessment do not harbor such faulty beliefs. While multiple-choice items can measure higher-order thinking and problem-solving skills it is certainly true that multiple-choice items cannot measure all possible outcomes. However, good item writers are able to write appropriate (e.g., tapping objectives beyond factual recall) multiple-choice items for mathematics, reading, and some portions of language arts curricula. (Writing should probably be assessed by asking students to write.)

Those who specify the proportion of items from different formats should be informed by measurement experts regarding which competencies can be assessed by which types of formats. State department officials need to recognize at the outset that it will be expensive, both in terms of time and money, to gather performance assessments on every high school graduate. Wainer and Thissen (1993) have found in their study of the Advanced Placement Chemistry Test that a 75 minute multiple-choice test in chemistry is as reliable as a 185 minute constructed response test. Because of scoring cost differences, the relative difference in costs for a given level of reliability is truly staggering. Wainer and Thissen (1993) estimated that if one, for example, wanted a test with a reliability of 0.92, it would cost 3000 times as much for a constructed response test as for a multiple-choice one.

Not only are there reliability and cost problems associated with performance assessments, but there are a myriad of other problems in areas such as validity, standard setting, and equating. Those non-measurement educators pushing for performance assessments should attempt to become educated with respect to these measurement problems. A recent survey of alternative assessments (Wolcott & Hoffman (1994) has concluded that "attaching high stakes to portfolio and performance assessment seems premature at this point" (p. vi). Certainly, if performance assessment (constructed response) exercises are used, there are a number of additional considerations regarding such issues as scoring, scaling, equating, and reporting that are addressed in later sections of this report.

Recommendation 16:
The department should provide, possibly through the technical advisory committee, statewide staff development for educators to increase awareness of measurement issues as well as the high cost associated with performance.

Recommendation 17:
Unless there is a compelling non-measurement reason, do not use the constructed response item format for competencies that can be assessed via multiple-choice items. Do not use any portfolio assessments (one type of performance assessment) for the MAAP.[9]

Finally, the State Department of Education must make a decision regarding how many items to develop initially. While this decision is related to other decisions (such as how many times a year to test, whether any given form can be reused, and whether anchor items are used for equating purposes), two general recommendations can be made.

Recommendation 18:
Contract for enough items initially so that after losses through pilot and field testing sufficient items will remain to build forms through the second administration year. There should be a longer range plan to develop a complete bank of items.

Recommedation 19:
Reissue a contract in sufficient time to have items developed and tried out (possibly embedded in a live form) prior to their being needed for the third year.

Pilot (field) Testing

We are very positively impressed with the recommendation of the High School Exit Assessment Implementation Committee that there be two years of pilot work on the new MAAP. However, as we

understand their current proposal for the first pilot, it would be to evaluate a variety of pilot assessment instruments from different vendors. While vendors will likely bid to produce a pilot, we are not convinced that this is the best way to proceed. One would have to write an RFP for the pilots, fund all those vendors who produce reasonable responses, test them all, go through some process to pick the correct vendor, etc. It seems preferable (certainly more efficient) to place an RFP that is as specific as possible for the "real" test and then Mississippi simply should choose the best response to that RFP. This would allow more focused attention from the very beginning on the test that will actually be used. Thus, the following recommendation.

Recommendation 20:
Issue one RFP for the development of the actual MAAP. Do not issue a separate RFP inviting vendors to build pilot tests from which you will choose "the best".

Another point that should be considered is whether some of the items for the new MAAP can be pilot tested through being embedded into the FLE. There are both positive and negative aspects to such a procedure. One positive aspect is that it should be cheaper. Another is that the students who take pilot items embedded into real tests will be motivated to try their best. A possible negative is that the items may be so different from those in the FLE that it will be apparent that they are pilot items. A second negative is that harder items on untaught content embedded in the FLE will negatively impact the morale (and therefore the performance) of the students who must pass the FLE. At any rate, some consideration should be given to piloting items within the FLE. The External Review Panel does *not* have a consensus recommendation regarding the wisdom of this. However the pilot testing is done, it is essential that enough items survive the pilot tests so that there will remain enough items for two years worth of actual forms. Depending on the frequency of test administration, this would mean enough items for four or six forms of the tests.

With respect to the proposed timeline, the HSEAIC has suggested the first pilot be in the spring of 1995 for all eligible students in grade 10, the second pilot to be in the fall of 1995 to obtain data for standard setting, and the first "real" administration to be to 10th graders in the fall of 1996. We have some concerns about this. First, we believe that it is very optimistic to expect the first pilot to be ready by the spring of 1995. Secondly, it would be preferable to have the pilot administered at the same time of year as the first real assessment will be administered. Thirdly, we think it would be nearly impossible to have a quality first pilot ready by the fall of 1995. Finally, as mentioned in the section on content specifications, we are concerned that the standards and competencies to be measured are not yet determined and that the process for developing the new curriculum structures for Reading and Language/Communication has hardly begun and is not scheduled for implementation until the 1997-1998 school year.

Recommendation 21:
Be flexible on the time lines. We believe it would be preferable to delay the first pilot until the fall of 1996, the second pilot in the fall of 1997, and the first real assessment in the fall of 1998. This means that the first graduating class affected would be the class of 2001.[10]

Following the pilot testing, the results should be disaggregated by appropriate demographic characteristics (e.g., gender, ethnicity, geographic location, students with limited English proficiency, students with disabilities, etc.). The results should be studied carefully for possible bias, quality of the items, etc. Careful consideration should be given to how widely the results of the pilot should be shared. If there are two pilots (as suggested by the implementation committee and as we strongly support) and the second pilot seems to be pretty much like the final assessment will be, we believe the results should be shared very thoroughly. This second pilot should ideally be viewed as a trial run of the actual implementation, rather than a repiloting of the items. This is a great opportunity to address all problem areas and program logistics. This would be an ideal time to try out some report forms to see if they are intelligible (user friendly). Further, because this assessment will cover more challenging content than the FLE it replaces, it is important that districts have an advance awareness of what will

likely be lower scores and, depending on the cut score, likely a higher failure rate, than they are used to seeing from the FLE.

Recommendation 22:
Consider what detail of reporting should follow the pilot tests. We believe the first pilot should be viewed as a combination research and development effort and results need not be widely shared. The results of the second pilot should be widely shared.

Recommendation 23:
Develop procedures and decision rules regarding which items in the pilots are OK, which need to be revised, and which need to be discarded. Develop documentation procedures regarding these decisions.

Scoring

The scoring of the objective portions of the examination should be contracted to a national scoring service. Commercial contractors have a great deal of experience and are well-equipped to do this scoring accurately and efficiently.

There are legitimate arguments both for having the performance assessments (constructed response items) scored by in-state teachers and for having them scored by an outside-the-state contractor. Reviewers of this document have made arguments on both sides of this issue. An argument for in-state teacher scoring is that teachers often enjoy and learn much from the scoring process. Using in-state teachers to score the papers may either add to or subtract from the credibility of the process, depending in part upon the quality of the training and monitoring process. At any rate, teachers should not be scoring papers from their own or surrounding districts as they could be aware of the identity of the students whose papers are being scored. If in-state teachers are to be used, the scoring sessions should be conducted by a scoring contractor with the clear understanding that teachers who cannot score reliably or validly will be dismissed.

Strong arguments can be made for using out-of-state personnel to score subjective tests. The major ones are timely scoring and costs. One state costed out the scoring of writing and found that using classroom teachers was the more expensive option. An "army" of teachers must leave their classrooms for at least four to six weeks two or three times a year (depending on how often the assessment is administered). These individuals must be paid their regular rates and substitutes must be provided. More important, however, their teaching expertise is lost during this time. Their students will never have the benefit of that lost instruction. Ways can be found to involve some in-state teachers without the disadvantages—for example, by using teams of teachers to observe the scoring process and using committees of teachers to assist in making policy decisions about scoring. We believe this preferable to actual in-state scoring. The scoring of high-stakes assessments in a reliable and valid manner is far more important than whatever staff development or public relations value there might be in using in-state scoring.

One could, of course, use both in-state and out-of-state scoring and compare the results. *The state should consider carefully the alternatives with respect to validity, credibility of results, costs, and ability to receive timely scores.* We offer the following recommendations.

Recommendation 24:
Contract for professional scoring.

Recommendation 25:
Develop a professional development packet to be used with teachers based upon the results of each year's scoring.

Standard Setting

When using a cut score on a test to determine whether individuals pass or fail, "the cut score becomes the linchpin in the decision process" (AERA, APA, NCME, 1985, p. 50). Yet, standard-setting is a subjective process, and typically there is dissonance between where policymakers think the cut score should be and the implication of that cut score for the failure rate (i.e., policymakers would typically think the cut score should be reasonably

high until they discover that such a cut score produces a "high" failure rate).

Much professional literature exists on the methodology for standard setting. In general, this measurement literature supports the following points: (1) A trained standard-setting committee should be involved in making recommendations regarding the standard. (2) This committee should use an iterative process that includes information about the failure rate by major ethnic groups (and perhaps other special populations). (3) The impact data *should be obtained from the first administration, not the pilot test.* (More needs to be said about this point. While the measurement literature would agree with this, and the panel members agree with this point when wearing their measurement "hats," there is a legitimate (non-measurement) argument on the other side. As one panel member has pointed out, there is a question regarding the practicality and timeliness of setting the standard *after* the first administration. From a practical standpoint, this puts the Department and State Board in the difficult position of administering a high stakes test and being unable to tell parents and students the score they will need to make to pass. Also at issue is how fast the students will need to receive the scores. The "wheels of policy-making" move slowly so that having the standard-setting committee convene after the test data are back and then making a recommendation to the Board could result in a considerable time lag. Thus, there may be considerable pressure to set the cut score before the first administration, and some states do that. However, other states do wait until real data are in, and measurement experts clearly prefer waiting.) (4) The recommendations from the standard-setting committee, a description of the process they used, a discussion of the relative costs of false positives and false negatives, and the fact that scores will go up across time should be taken to the group officially responsible for setting the standard, and this group should make the final decision regarding where to set the cut score. On high-stakes tests where the content of the tests is at a reasonably high level (as on the MAAP), it would generally be considered inappropriate to simply set the standard at 70%—as we believe was done on the FLE. However, the philosophical concept of allowing for a lower score on one test to be partially compensated for by a higher score on another test—as is being done on the FLE—is appropriate from a technical point of view if it is philosophically/educationally acceptable. If one does employ a partial compensatory model, the members of the various standard setting committees should be aware of this and they should be trained accordingly.[11] We do not give specific recommendations about the training process, but rather, recommend further advice about this difficult problem.

The following broad recommendations are made regarding standard setting.

Recommendation 26:
Reconsider the current plan to set the standard using the pilot study results as impact data. While this may be preferable from some practical points of view, it is not the approach preferred by measurement experts.

Recommendation 27:
Appoint and train a standard-setting committee. This committee should be composed of individuals who are both qualified and credible. A majority of the committee probably should be Mississippi public school educators with knowledge and experience both in the subject matter being assessed and at the grade level of the students being assessed.[12]

Recommendation 28:
Use a technical advisory committee to help develop a specific standard setting procedure.[13]

Recommendation 29:
The State Board of Education should establish a passing score through administrative rule based upon a recommendation by the superintendent of public instruction with the advice of appropriate committees.

Setting standards for performance assessments has been considerably less researched than setting standards for multiple-choice tests. One could, of course, set separate standards for the two item formats within a subject matter and use a conjunctive model for making the pass decisions (i.e., one would have to pass both "subtests" within the subject matter). Setting separate standards for the

two formats is not recommended by us for a variety of psychometric reasons. In the first place, it is unlikely that the performance assessment portion of the assessment would be long enough for one to place any confidence in either its reliability or validity as a stand-alone assessment. Second, such a process would make even more difficult the equating problems. Nevertheless, combining both types of formats poses formidable problems. Given the state of the art, it is reasonable to suggest the following recommendation.

Recommendation 30:
Engage in several small scale pilot study approaches to setting standards on assessments composed of multiple-choice and performance assessments. Do statistical analyses regarding the impact of these approaches.

Because the initial failure rate probably will be greater than the failure rate after the test has been in place for several years, it may be reasonable to set incremental cut scores over time. This allows the cut score to be set so that an inordinate number of students do not fail at the beginning, but the state is not locked into a cut score that is lower than desirable. The advantage of setting these incremental cut scores at the beginning is that it may be easier to do than to reset the cut scores later.

Recommendation 31:
Consider setting incremental cut scores for different graduating classes when the State Board of Education makes its initial decision.

Another issue regarding standard setting is what standard should be set for accreditation purposes. We address the issue of using various pieces of test data for accreditation in more detail later in the report. However, it seems worth mentioning here that if one wants schools to strive to get all students above the cut score so that they can graduate, any accreditation use rule should not force districts to choose between working for accreditation and working for a high (e.g., 100%) pass rate for the students.

Item Sensitivity Reviews and Empirical Bias Studies

All assessments should be designed to be free of ethnic, cultural, and gender "bias." There are well-developed methods to eliminate such bias. The first is in the training of the item writers. They should be trained to avoid certain stereotypical words and phrases that may be offensive or may give an unfair advantage to a particular ethnic, cultural, or gender group. (Another group that should be considered is the Vocational Technical students. One individual we interviewed was concerned about the fairness of writing prompts for those individuals. This is a legitimate concern and should be kept in mind. It would also be appropriate to have the item writers keep in mind other special populations such as those who have certain disabilities.)

A second procedure is to have all items reviewed by a committee of individuals specifically trained to detect items that may show such insensitivity. The item sensitivity review team can be trained to focus on a variety of different groups such as those discussed in the previous paragraph.

A third procedure is to compute "differential item functioning" statistics on all of the items based on a pilot study (field tryout). Due to the numbers of individuals that exist in the different groups, these statistical procedures can probably be done only on major groups such as both genders and the predominant ethnic groups. Those items that are "flagged" by such a statistical analysis should then be brought back to the item sensitivity review committee—and probably to the relevant subject matter content committee—for a final determination of whether those items should be removed from the item bank. A fourth procedure is to collect committee members' judgments on whether or not the test as a whole is relatively free of bias.

It should be pointed out that when a test is composed of items with different formats and those items may carry different "weights" the task of scaling the test (see below)—which is a prerequisite to empirical differential item functioning procedures—is a bit more intricate, and prospective contractors should respond to an RFP with details regarding how they will proceed with such a task.

It is important to note that while the test should be free from "bias," this does not mean that all ethnic, cultural, and gender subgroups should necessarily have the same mean level of performance. If some groups truly have not achieved as many of the skills in one of the subject matter areas (or indeed on a particular item), the test (item) should reflect that true state of affairs. Based on the findings from many previous assessments, the Mississippi Department of Education should anticipate that not all subgroups are achieving at the same level and that the test scores will show those differences. The purpose of the item sensitivity reviews and the differential item functioning studies is to gather data to allow for informed judgments about whether the individual items and/or the test items collectively contain irrelevant content that results in unfairness to a subgroup.

Recommendation 32:
The item sensitivity reviews should be completed by a committee that is selected and trained specifically for this task.[14] Most members should represent the state's predominant minority groups. However, it would be wise to include at least one member of the committee who is a minority group member from out-of-state and a recognized expert in this area.

Recommendation 33:
Conduct statistical differntial item functioning differentially for different groups should be flagged and reviewed (but not necessarily discarded) by an item sensitivity review committee (conceivably-but not necessarily-the committee used for the item sensitivity review) and a content review committee. Clear guidlines should be developed regarding how to respond to flagged items, how to handle committee members' disagreements, etc.

Reliability

Reliability pertains to the amount of test variance that is due to random error. Data should have high reliability. There is currently some debate about just how reliability should best be calculated for performance assessments. The state may want to obtain specific advice from a technical advisory committee on the best way to combine performance and multiple-choice assessments and whether to obtain reliability through a battery reliability formula or some other approach. While those responsible for monitoring the quality of the assessment should study various approaches and ask the contractor in an RFP to provide specific recommendations, we offer the following recommendation.

Recommendation 34:
Obtain the following reliability estimates; internal consistency, interrater reliability[15], generalizability across performance samples, and the reliability or standard error at the cut score.

Scaling/Reporting

Once tests have been scored, the students' results must be reported. Generally, it is not considered wise to report the "raw scores" (e.g., number of items right on a test). The scores are typically reported based on some mathematical transformation of the raw scores so that the transformed scores have certain statistical properties (e.g., a specific mean and standard deviation). When multiple-choice items and performance assessment items are combined into the same test and one wants a single score, there are difficult decisions to be made regarding how to combine the two sets of items. The easiest approach (*not* necessarily the best) is to determine in advance how many raw score points to assign each level of performance on the performance assessment items and to simply add these points to the number of multiple-choice items an individual gets correct. A second approach would be to score the two sets of items separately and then combine those two scores through some *a priori* weighting scheme. That scheme could be based on a logical/philosophical weighting or an empirical weighting based on any number of different variables such as their separate reliability estimates, subtest information functions, etc. A third approach would be to use an IRT model that scores all the items together. There, the typical choices are a one-parameter or a multiple-parameter model. Another approach is not to combine the two types of assessments at all, but to scale them separately and to set two separate cut scores. How the combining gets done (or whether the scores are kept

separate) has relevance for reporting and for equating approaches considered later. How the weighting/combining gets done also could be extremely important to an individual student. In addition, a particular combination method may be more beneficial to some subgroup (e.g., one gender or some ethnic group).

Recommendation 35:
Consider carefully how the performance assessments and multiple-choice items are to be combined. There should be expert advice regarding this and empirical studies showing the differential impacts of various approaches on individual students, groups of students, ease and quality of equating procedures, etc.

Because high school graduation tests are not typically designed to differentiate among those passing, and because one should not encourage use of information on the difference in students' scores above the cut score (e.g., for employment decisions after graduation or for district accreditation decisions), one would typically report scores above the cut score only as a "pass."

Other questions arise for those who do not pass. Educators typically want high school graduation tests to be diagnostic. They believe that failing students should be given some information that would facilitate efficient and effective remediation efforts. That is understandable, but it is difficult to design a test that is of high quality both for determining accurately who deserves to pass and for determining just what the specific diagnostic recommendations should be for individuals who fail. Thus the dilemma. Reporting sub-test scores may imply more diagnostic information than can be justified based on such technical considerations as the reliability of the difference scores. However, not to report sub-test scores limits the usefulness of the scores for remediation. Because reporting sub-test scores is a multifaceted and technical issue, it deserves careful attention.

If the decision is made to report sub-test scores (as we suspect it will be), it will have implications for the test specifications and test development. Troublesome problems may arise if the assessments make use of a variety of item types. The subscales composed of primarily performance assessment exercises are not likely to be as equivalent across years as the subscales composed of multiple-choice items. This may have implications for how to communicate the subscales. Further, the item weighting of the constructed response and multiple-choice items will impact the decisions regarding subscale reporting.

Careful thought should be given to what has been learned regarding how the subscores reported for the FLE have been interpreted and used. Personnel from the State Department of Education should make concerted efforts to determine the acceptability (both from a public acceptability point of view and from a psychometric point of view) of the current approach and determine whether changes need to be made for the MAAP.

The issue of which transformed scores (scaled scores) to use for reporting is also a difficult technical issue that cannot be solved in the abstract. Numerous scores could be used. Using the same scaled scores across subject matters does have some advantages, and we would recommend it. However, using a common scale across subject areas may have implications for test development.

Recommendation 36:
Scores should be reported as pass or fail. Those individuals who fail should be given some information regarding how close they were to passing, and they should be given some diagnostic information that would facilitate remediation efforts. Important technical details (e.g., reliability of difference scores) regarding various methods of reporting diagnostic information should be worked out and specific plans should be formulated by a technical advisory committee prior to approval of the final test specifications.

Recommendation 37:
Use a common scale across subject matter areas. This takes some advance planning to avoid adopting a scale that is appropriate for one test but unworkable for another.

Recommendation 38:
Consider whether it would be better to keep the same scaled score approach as is being used on the FLE or whether it might be better to change the score to avoid confusing the two.[16]

If the assessment were to be norm-referenced, there should be a reporting of an individual's norm-referenced score. Although we realize that there has been some discussion in Mississippi of the MAAP being norm-referenced as well as criterion-referenced, we see extremely troublesome measurement and legal problems in trying to develop an exit assessment that both matches the Mississippi curriculum and is normed on a representative national sample. We would much prefer to have the MDE use the norm-referenced ninth grade *TAP* or the *Work Keys* test as a measure of how Mississippi students are doing in comparison to a national average than to try and make the exit examination both nationally norm-referenced yet have the content be representative of state curriculum structures.

Number of Forms

The number of forms that need to be available for the MAAP deserves careful consideration. (We are not considering forms that are identical except for the pilot items as being separate forms. It should also be pointed out that the number of forms available for the *subject matter tests* and for *Work Keys* are relevant issues for consideration.) We offer the following recommendation regarding forms.

Recommendation 39:
Develop rules/procedures for designing forms for makeup examinations and out-of-school (i.e., adult education) populations.[17] Determine whether forms will be reused. Determine how many times you will administer the test each year. Determine equating procedures (e.g., number of anchor items). Based on these considerations, develop enough alternate forms to last through the second year of test administration. Develop more forms/items during this time so that a sufficient supply is continuously available.

Equating

High school graduation test questions need to remain secure, and they cannot be reused to any great extent. (We would make the same statement about the course exams and Work Keys tests.) However, to be fair to individuals who take different forms of the test, the forms need to be equated. It is particularly important that diploma-sanction tests be equated *at the cut score*, so that a performance level that was considered a pass on one form of the test would not be considered a fail on a different form. There are many ways to equate, but we should stress that the process becomes a bit more difficult if there are both performance assessment items and multiple-choice items. The process becomes even more difficult if the proportion of the item types does not remain constant across forms and/or if the decision has been made to scale the two item forms separately.

There are many ways to equate, but the two more common general procedures considered viable for diploma sanction tests are to use anchor items or to pre-equate. Anchor item equating is generally preferable to pre-equating for final cut score decisions, because the subareas of the test will likely be differently affected by instructional changes. Pre-equating should be done when initially building various test forms. The cut score will, of course, be set on the original form. The wording of the rule adopting a cut score needs to be carefully considered so that it is clear how to equate that score to scores on subsequent forms of the test.

Recommendation 40:
Use a technical advisory committee to help develop specific equating procedures.

Standardization of Test Administration

Recommendation 41:
Carefully consider policies regarding all test administration conditions. For example, the decision whether or not to use calculations in the mathematics test must be constant across all administrative sites. Train personnel adequately to administer the tests. Consider random auditing of the administration process to ensure uniformity throughout the state.

Education Issues

ALL OF THE ISSUES INVOLVED IN A high school graduation test (as well as the other Mississippi assessments) could be considered educational issues. However, in this section, five special kinds will be discussed: articulation with other tests, retesting, remediation, special education, and adult education.

Articulation with Other Tests

The Mississippi high school graduation test should be articulated with the other tests in Mississippi. (This seems like an appropriate time to stress that there needs to be articulation among the various committees (e.g., norm-referenced committee, exit test implementation committee, etc.). Several individuals we interviewed thought the communication across these committees was less than ideal.) Of particular concern should be the early testing (e.g., the *ITBS*, the *Performance Assessments*, and the *TAP* given in grades 4–9), the *end of course tests*, and the *Work Keys* tests. As we understand the current accreditation procedure, the various tests all count in a formula for determining accreditation. Thought needs to be given as to whether these various tests and their specific uses within the state complement each other or result in competing goals.

In a subsequent section on accreditation we will discuss some of these issues further. Here we would like to point out that, whatever the uses of other tests in an accreditation system, Mississippi should administer tests in earlier grades that would assist in identifying students who may not be acquiring prerequisite knowledge and skills at the expected rate to enable them to pass the MAAP. Attention needs to be given to the relationship of the content on the grades 4–9 tests to the content that is on the MAAP. Ideally the early grade testing would be testing for the specific prerequisite knowledge and skills that are important for passing the MAAP. If not, the early tests could not be used to identify those likely to need additional instructional support prior to taking the MAAP. While ideally, there would be a relationship between the contents, it seems important to call the reader's attention to some additional concerns. It is surely possible for a student not to have acquired some prerequisite knowledge and skills by, say, grade 8, yet that student—with appropriate effort—may well acquire the knowledge and skills necessary to pass the MAAP. Likewise, doing well on an 8th grade test that covers prerequisite outcome measures in no way guarantees that a student will acquire the outcome measures sufficient to pass the MAAP. This latter point needs to be made very clear to all students, parents, and educators. Early tests should not and will not cover all the competencies assessed on the MAAP.

Recommendation 42:
Have subject matter experts study the content of the grades 4–9 tests and the competencies to be measured on the MAAP. If appropriate content articulation does not exist, determine whether the problem should be fixed by changing the content of the early tests or the MAAP.

Recommendation 43:
Even if close content articulation exists, be cautious about any "predictive" interpretation of the scores of a single individual from testing in earlier grades. Such tests should be thought of as providing only an early awareness, not a strong, reliable predictor.

Consideration should also be given to whether the course subject matter tests assess the same competencies as the MAAP. For example, if the content of the Algebra I examination covers many of the competencies assessed on the MAAP, and if a student passes the Algebra I exam prior to grade 10, consideration should be given as to whether it should still be necessary to pass the mathematics portion of the MAAP.

Recommendation 44:
Consider whether passing any of the course tests can serve as alternatives to passing certain of the MAAP tests.

Finally, there should be articulation with the *Work Keys* tests. If the content of those tests are quite different, then we have the same types of articulation concerns as would exist if the other assessments differ from the exit examinations.

As has probably been apparent from the previous discussion, there needs to be concern with the total amount of testing as well as the articulation of the tests. It is certainly possible that there is simply too much testing being planned at the high school level and we believe educators in Mississippi need to review each projected program and assess the purpose of it.

Retesting

Retest issues are of two types: how and whether to give makeup tests for absentees (not a retest of the same person), and how many chances a single individual should have to pass the test.

If someone is ill or has an excused absence on the day of a test, that person should have an opportunity to make up the test as soon as possible. The state must consider whether the district/building should have a window of opportunity in which it can retain the tests and provide an opportunity for makeup tests. This provision seems appropriate if the window of opportunity is not too long; we suggest approximately one week total. Special consideration should be given to the issue of whether alternative forms of the writing prompts and the performance assessment portions of the other assessments need to be used for makeup examinations. Extended absences should be handled on a different basis. Written policies should be formulated regarding all makeup procedures.

Other retake issues include the following: Is the student who fails a test area (e.g., writing) required only to retake the failed area; is a student who fails the test obligated to retake that test during each succeeding administration or may the student "sit out"; and when a school is closed by a crisis, can the test administration be rescheduled for that particular school outside of the announced "window"?

Recommendation 45:
The department should prepare and the board ahould adopt specific written procedures regarding makeup examination provisions.

The number of permissible retakes also should be a matter of policy. Evidence in other states suggests that four or five total attempts prior to scheduled graduation should be sufficient. A person should be allowed free, unlimited retakes through an adult education program if the person has not passed during the regular high school time period.

Recommendation 46:
The department should prepare and the board should adopt specific written rules regarding the number of retakes that should be allowed and how many attempts a student should be given prior to the time he/she is scheduled to graduate.

Remediation

We are aware that the Mississippi Department of Education is interested in pursuing its role in advancing the professional development of teachers. This is commendable with or without high stakes exit examinations. When a state requires that students acquire certain competencies (as measured on an exit exam) prior to graduating, that state should have some responsibility for assisting the local schools in planning for remediation. It seems wise that a state rule should be established to provide that a child who fails must be given the opportunity for remediation.[18]

Several issues need to be considered regarding remediation. For example, who is responsible for designing remediation materials—the local school or the state? If the state designs the materials, is it responsible for evaluating the materials for their effectiveness? Should the state hold workshops around the state on how to remediate? Should the state attempt to control the publication of materials by commercial publishers? If remediation programs increase the costs to the local districts, will they be reimbursed by the state? How can remediation be completed without the negative side effects of tracking or grouping? If a student who has not passed the graduation test requirements but has passed all other requirements decides to return to school for a 13th year, can that student by counted for state aid? Will local schools be required to document their offers of remediation to those who fail?

Recommendation 47:
Develop a detailed proposal (set of guidelines)

that addresses questions regarding remediation efforts and the respective responsibilities of the state, the district, and the student for remediation efforts.[19] This set of guidelines should then be approved by the department and the state board.

Legal Issues

ANY HIGH SCHOOL GRADUATION test should be built so that it is technically sound. Furthermore, decisions made from the data should be applied fairly. Generally speaking, if one can provide evidence regarding those issues, the process should be legally defensible. Thus, we have already addressed legal issues and will continue to do so in sections following this one. However, some more specific legal issues should be kept in mind and are addressed in this section.

First, the state should be aware that tests are frequently questioned from a technical standpoint. The courts will use the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985). [It should be pointed out that the process of revising these standards is underway, and readers must remain alert to what the new standards say when they are published—probably not before 1996.] With respect to legal issues, it is wise to obtain legal involvement early from the attorney general's office. This may be less urgent for Mississippi because they already have an exit examination in place. However, it would be our expectation that the failure rate for the new examination may be considerably greater than what it has been for the FLE. Further, there is discussion of using the MAAP to force curricular change—a tactic that is likely to meet legal challenge.

Liability Issues

A thorough investigation of liability issues should be made. Do existing state statutes protect employees? If the state department retains the service of local educators, does any state statute protect them? Can a teacher be sued because of a claim that he/she did not teach some content—or teach it well enough? Are committee members who make recommendations covered under state statutes?

Recommendation 48:
Carefully investigate liability issues with assistance from the attorney generals office. Attempt to obtain necessary statutes with respect to liability. Inform all committees and all staff regarding their potential liability.

Notification

One of the main legal issues other than test quality is due process. Individuals need sufficient notification of the *new* graduation requirement. This notification should be detailed with respect to the standards and competencies that the tests will cover. Details concerning how to notify students and parents need to be worked out. Certified letters need not be sent to every child/parent. Nevertheless, there should be some documentation that the notices were sent (announced). Procedures such as placing notices in a student handbook, placing notices on report cards, etc., should be considered. One suggestion is to produce a video tape to show all students and have each district provide an affidavit that they have shown the tape to all ninth graders. Whatever is done regarding notification for the first cohort should be continued for all future classes.

Recommendation 49:
Schools should be notified immediately regarding the NEW graduation requirement and the information disseminated to all teachers. Students and their parents should be notified no later than the year in which affected students are in the ninth grade. The public in general should be notified immediately following decisions made by the state.[20]

Timing

As mentioned above, due process requires *sufficient* notification. Thus, the amount of lead time becomes an important legal (and educational) issue. As we discussed previously (see Recommendation 21), the exam may indeed not be developed in time for the necessary piloting and revisions so that it can be used for a graduation requirement prior to the graduating class of 2001. Whether or not the test could actually be ready, there is the issue of sufficient due process. A general rule of thumb

might be that the students who are to be impacted by the assessment be notified of the specific standards and competencies to be assessed no later than when they are in ninth grade. This would mean that notification would need to be given no later than the fall of 1997 if our suggested guideline is followed.[21] Given that the estimated implementation date for the new English/Language Arts and Reading Frameworks is 1997-1998. This seems like reasonable timing.

Related to the timing issue is when to phase out the FLE. If the new MAAP is to be required for the graduating class of 2001, the FLE would need to be administered through the spring of 2000.

Recommendation 50:
The FLE should not be used for accreditation purposes after the first year the MAPP is used for such.[22]

Opportunity to Learn

As we have discussed previously, it is illegal to require students to pass a test that covers standards and competencies unless it can be shown that the students have had the opportunity to learn that set of material. It would be inappropriate to require the new exams for graduation until it could be demonstrated that the new curriculums were in place in the districts, that the teachers had received sufficient professional development so that they knew how to effectively teach the new curriculum, and that, indeed, the students had an opportunity to learn the new material. We remind our readers of an earlier recommendation that if it cannot be shown that students have had an opportunity to learn the new curriculum, the assessment should be postponed.

Documentation

The general issue of documentation also needs some attention.[23] The lack of various types of documentation can become a central focus of a law suit. We are not totally aware of the documentation policies for the FLE. However, these should be reviewed to determine whether they are sufficiently detailed. For example, when committees review items for sensitivity or bias, consideration should be given as to whether a complete record should be kept regarding which individuals considered which items biased and what changes to the items resulted if they were revised.[24] One also needs to consider how long any documentation should be kept.

Recommendation 51:
If sufficient documentation policies do not exist for the FLE, the department should prepare, and the board should adopt, detailed policies regarding what should be documented and how long the documentation should be kept on file. A general suggestion is that all documentation be kept for a period of atleast five years following the school year in which the test was administered. Consider keeping "forever" the initial development documentation and records about when, why, and how procedures are adopted and/or changed.

Security Provisions

We are aware that Section 37-16-4 of the current Code regarding the Statewide Testing Program discusses violations of test security procedures and penalties. However, we believe the department should consider whether there needs to be additional statements regarding what constitutes inappropriate, unethical, unprofessional, and possibly illegal behavior on the part of educators and students with respect to violating administrative standards, security procedures, and so forth.

Recommendation 52:
In consultation with the attorney general's office, the department should prepare and the State Board of Education should adopt rules on what constitutes inappropriate behavior on the part of educators or students with respect to test taking, security issues, and so forth, and what penalties will be imposed for violation of these rules. These rules and the penalties should be disseminated to educators, students, and parents prior to the initial administration of the MAAP.

Recommendation 53:
Test security provisions must be a shared responsibility among the contractor for test administration, the state department, and the local schools.

Accommodations

As mentioned earlier in the discussion prior to Recommendation 1, attention needs to be given to whether the current Code and other policies are sufficient regarding accommodation practices.

Recommendation 54:
Review accomodation codes/regulations to determine whether they need to be updated.

Policy/ Administrative Issues

A PLETHORA OF POLICY/ADMINISTRATIVE decisions must be made and rules must be passed prior to implementing a high school graduation test requirement. Obviously the State Department already has made many of the necessary decisions and rules because they would be much the same for the FLE and the proposed MAAP. We list below a set of questions that, if they have not already been answered, will need to be considered by the Department.

- Who approves the various test construction and test administrative procedures?
- Who develops, approves, and oversees all test security issues? Is there a procedure in place to monitor the districts to assure they do not issue diplomas to those who have not passed the MAAP?
- Are there sufficient equipment/facilities for storage of secure materials, shredding out-of-date secure materials, and so on?
- Has it been determined how to handle retakes for those who have completed all other high school requirements and have "left" school? Has a policy been established for issuing diplomas to adults?
- Do all transfer students from other states (even those transferring during the second semester of their senior year) need to pass the MAAP to receive a diploma? What if those students have passed another state's graduation test in the same subject? How about students who transfer within the state from non-public to public schools?
- What, if any, accommodations or exemptions will be permitted for students on an IEP or 504 Plan? What about those whose language spoken in the home is not English, migrant students who move in and out of the state, or those who are simply foreign exchange students spending less than two years in the state? What is the intent with respect to language of the exams? Is it the intent that all tests should be in English, or only that students should read and write English?
- What happens to a senior in the year prior to the effective date of this graduation requirement who fails a required second semester course that must be completed in summer school or in the first semester of the next year (when the graduation requirement applies)? Will such a student also have to pass the test even though he originally was not required to do so?
- Will the state have a policy on participation in commencement exercises by students who complete curriculum requirements, but not test requirements? Will such a student receive anything—e.g., a certificate of attendance or a document verifying accomplishments?
- Who approves various external committee appointments? Should there be written policies regarding representation on those committees?
- Who finally sets or approves the cut score following the recommendation from a cut score committee? Will the respective costs of false positives and false negatives be considered and, if so, by whom?
- Is the system of tracking students that is being used for the FLE working ok? If not, how does it need to be changed?
- Are the reports being used for the FLE sufficiently detailed? Should there be more attention to studying the results via such procedures

as disaggregating the results by ethnicity, courses taken, etc? Should the results for the separate item formats (e.g., multiple-choice and performance assessments) be reported separately and should there be the same disaggregation of the results by item format as for the total results?

- Who will develop the total annual test administration plan and how will it be communicated to all school districts? Will test administrations be monitored by the state?

Recommendation 55:
Consider the questions such as those raised above and make the necessary decisions concerning them. The department and the Board of Education must devote adequate time to the identification and resolution of critical questions that must be addressed.

Human and Financial Resource Issues

LEGISLATORS CANNOT BE EXPECTED to recognize the huge additional costs of implementing a high school graduation assessment that is composed of both multiple-choice and performance assessment exercises. The State Department of Education must provide a rationale to them to support any request for additional human and financial resources. This section discusses needs in staffing, advisory committees, contractors, and financial resources.

Staffing Needs

It is our understanding that the Student Assessment unit has a professional staff of either five or six individuals—counting the director. (Our notes suggest five total but the *Summary Report for 1994 Mississippi Statewide Testing Program* lists six individuals in the Office of Student Assessment.) Given the total assessment program for which this staff has responsibility, we believe it is important to increase the size of the staff. Even though a large portion of the work will be contracted out to various vendors, there remains a great deal of additional work that must be done by staff. For example, an individual should be assigned major responsibility for each content area to be assessed. A measurement specialist with technical background will need to spend considerable time writing RFPs. Specific tasks for the contractors need to be developed and the contractors' execution of these tasks needs to be monitored. Someone must coordinate the assessment staff in the areas of test development, test administration, and test use and reporting. There needs to be an overall supervisor.

There may also need to be additional staff in the curriculum/instructional area. We are not aware of just how many professionals are employed in these areas, but with the advent of the revised curriculum structures and the new assessment, much professional development of staff in the districts needs to take place.

Recommendation 56:
The department should conduct a careful study to access additional staffing needs in the student assessment and curriculum/instructional units. We should think that, at a minimum, the new set of assessment plans would call for some additional professional staff in the student assessment unit. There probably needs to be additional staff added in the curriculum/instructional unit as well.

Advisory Committees

The need for several advisory committees has already been discussed in various places in this report, and further information about our recommendations regarding the composition of these committees can be found in the next section. However, for the ease of individuals interested in human and financial resource needs, they are listed here under a specific recommendation.

Recommendation 57:
If they have not already been established, the following advisory committees should be appointed: A testing policy advisory committee, an item sensitivity review committee, a technical advisory committee, a content review committee

in each content area of the assessment, and a committee to recommend a standard (cut-score) (one such committee for each subject assessed).[25]

Contractors

Mississippi has considerable experience dealing with contractors, and we suspect they have done an admirable job. While we are unaware of state restrictions on contracting procedures, we hope the Department has, or will be given, the freedom to grant single source contracts and to issue agreements that extend across fiscal years. We have one recommendation that is based on considerable experience that it is advantageous to keep the number of contractors down to a reasonably small number.

Recommendation 58:
For the MAAP, the department should consider using at most two contractors: one for test development and formal field tryouts and another for test administration, scoring, and reporting.[26]

Financial Resources

The need for appropriate staff, advisory committees, and outside contractors relates to financial needs. The specific costs depend on decisions regarding many of the issues already discussed in this report. Costs under some test designs easily can be more than triple what they would be under other designs. For example, the higher the proportion of the assessments that are performance-based, the higher will be the costs of administering and scoring the assessments. Two specific issues that have not been considered earlier and may have cost implications are (1) whether non-public students will be tested (even though they are not required to pass to receive a diploma) and, if so, who will pay the cost, and (2) whether the state is responsible for the financing of state-required local school functions (e.g., professional development of staff and costs of local administration of the assessments). Other states can provide detailed information about various costs, and we urge Mississippi personnel to contact them. For example, Florida and Ohio have been using multiple-choice tests for high school graduation for years, and directors such as Tom Fisher (Florida) and Roger Trent (Ohio) would be able to provide estimates for what their states are paying. Louisiana has a program with many similarities to the proposed Mississippi program and Rebecca Christian and her staff would be a useful resource. Michigan is in the process of having a high school exit test developed (for state endorsement rather than diploma purposes but that would not affect costs) that will include performance assessment exercises. Diane Smolen could provide information regarding the development costs for Michigan assessments. Many other states also have high school exit tests, and it is our experience that the directors are very willing to assist other state directors by providing information regarding costs of their programs.

Recommendation 59:
Obtain information from other states with similar programs regarding fiscal needs. Make recommendations to the legislature that are sufficient to cover department needs, and make clear to them that the task simply cannot be accomplished without adequate support.

SECTION III:

Sequencing of Tasks

IN DESIGNING A PROGRAM FOR A HIGH school graduation test, it is useful to have in mind the total set of processes and approximate completion dates for various activities. While we recognize that Mississippi already has considerable experience in designing and implementing a high school graduation test (the FLE), it might be useful to list, in abbreviated fashion, the tasks we believe are required and some suggested timelines. The timelines are based on the assumption that our recommendation regarding the new test impacting the graduating class of 2001 is followed. Obviously, the suggested sequence and timelines are based on certain assumptions about decisions reached. Different decisions would result in different steps/timelines.

It is important to note that many process strands actually run concurrently. Furthermore, missing

one or more of the targeted deadlines can mean that all other deadlines following that one are missed and that the program cannot be implemented on time. Both the legislators and the Board of Education need to understand that a lot of work needs to be done and that it takes sufficient staff and resources to accomplish the tasks.

Below is one possible sequence of activities that could be carried out to develop and implement the MAAP. It represents a sequence that we believe to be a reasonable approach. Detailed suggestions about how to perform those activities are not present in this section. The text and recommendations in the previous sections cover many such details.

Sample Tasks and Completion Dates

(assuming requirements are for the 2001 graduating class)

Task 1: Establish appropriate advisory committees. Do this as soon as possible. This task involves determining what committees need to be established, determining criteria for selection of the committee members, soliciting and evaluating the nominations, officially appointing and training the committee members, and maintaining the committees over time. We suggest the following committee with the understanding that it might be wise to have some overlap of committee members:

- **Department of Education Steering Committee:** This committee should represent the various units of the Department whose tasks will be impacted by this program (e.g., the Student Assessment, Curriculum, Vocational, Adult, and Special Education units).
- **Testing Policy Advisory Committee:** This committee would be much like the previous task force or the current implementation committee. It should represent the state education community to advise on policy.
- **Item Sensitivity Review Committee:** This committee should be composed mostly of the state's minority groups, but with at least one member from out of state who is a recognized expert on bias issues in assessment.
- **Technical Advisory Committee:** This committee should be composed of at least one measurement expert from within the state and at least one individual who has been (or is) the director of a similar competency testing program in another state. Other members of the committee should be widely recognized as measurement experts, and they (as a group) should have expertise in test development, scaling, equating, and all other major areas about which the department may wish to obtain advice.
- **Content Review Committees:** These committees should be composed of content experts (mostly or totally state residents) in each area of the test. State department personnel who are specialists in the respective subject matter areas should sit on these committees, although it is debatable whether they should have the right to vote.
- **Standard-setting Committees:** These committees should be composed primarily of individuals who are both qualified and credible. They should probably be primarily composed of educators in the state who have knowledge and experience both in the subject matter being assessed and at the grade level of the students being assessed.

Task 2: Determine exactly what standards and competencies will be assessed. As we suggested earlier in the report, this is a very important task and one we believe is far from complete. Once this is determined, there should probably be specific approval of those standards and competencies by the State Board of Education. Ideally, this task should be completed no later than May, 1995.

Task 3: Disseminate information about Task 2 to all students who will be impacted, parents, business leaders, and other relevant constituencies. Complete before schools let out for the summer of 1995.

Task 4: Complete test specifications for each test area. Complete by August, 1995.

Task 5: Hire a contractor for development of item specifications, item/test development, and field tryouts. Complete by December, 1995.

Task 6: Have the contractor complete the item specifications, item writing, informal pilot testing, and item editing. Complete by February, 1996.

Task 7: Perform content committee review and revisions as necessary. Complete by March, 1996.

Task 8: Produce camera-ready copy for formal field tryouts. Complete by July, 1996.

Task 9: Field test items first time in Fall on students in Grade 10. Complete by early Fall, 1996.

Task 10: Prepare and disseminate descriptive information and sample test items to assist in preparing teachers, students, and parents. Complete by early Fall, 1996.

Task 11: Develop and adopt rules governing test administration, scoring, and reporting. Complete by Spring, 1997.

Task 12: Analyze field test and revise items as necessary for second field test. Complete by Spring, 1997.

Task 13: Conduct second field test. Complete by early Fall, 1997.

Task 14: Revise items from second field test as necessary. Select items for the required number of forms needed for the first year of the real testing from the subset of items that did not need any substantive revision. Complete by early Spring, 1998.

Task 15: Select operations contractor for administration, scoring, and reporting. Complete by early Spring, 1998.

Task 16: Conduct regional seminars for school administrators and testing coordinators on the administration, scoring, and reporting procedures. Complete by early Spring, 1998.

Task 17: Complete production of all necessary materials for first tests and have them ready for distribution. Complete by Summer, 1998.

Task 18: Administer first real test to tenth graders (class of 2001). Complete in Fall, 1998.

Task 19: Score, analyze results of first administration, and establish passing standards for the first administration. Complete in late Fall, 1998. (As mentioned earlier, the panel recognizes that a practical case can be made for establishing the standards after the second pilot.)

Task 20: Design and implement a plan for releasing test results to the schools and the general public. Complete in late Fall, 1998.

Task 21: Review and repeat steps above. Plan extended timeline to include at least two administrations per year for 10th through 12th graders. Include time for equating procedures for future test administrations. This task should be carried out continuously.

Section IV:

Using Test Scores for Accreditation Purposes

WE HAVE BEEN GIVEN COPIES OF THE "Accreditation Requirements of the State Board of Education" (Bulletin 171), and have had the opportunity to meet with individuals in Mississippi who are considering revisions (refinements) to these requirements. The new assessment program, including a revised exit exam and other additional testing in the schools, should be considered while making these revisions.

One issue has to do with what standard to use in setting performance standards for all the measures. Previously, the accreditation standards on the FLE and the Stanford Achievement Test used average scaled scores. Holding districts accountable for raising average scaled scores may provide incentives for the district that are incompatible with the purpose of the exit test, and indeed, may be incompatible with how some educators would like to see resources expended for achieving the standard on the elementary school achievement tests. It may be easier to raise average scaled scores by concentrating instructional attention on those that already score above the standard for graduation. That is, schools may be rewarded for allocating resources primarily to assist individuals who would have passed the MAAP on the first attempt anyway instead of providing help to lower achieving students who are at risk of not graduating.

Recommendation 60:
In the accreditation system, the "success of the school system" could and perhaps should be defined in terms of the number of students who demonstrate the desired level of performance rather than in terms of average scores.[27] In any case, to maintain the intergrity of purpose for the MAAP, the satndards at least for that exam should relate to the proportions of students who are successful on a specified attempt.

Assuming the above recommendation is followed, another issue to consider is whether schools should be held accountable for the percent of students who pass the MAAP on the first try or the cumulative percent who have passed on some future attempt. Individuals attain desired levels of achievement at different rates. Some individuals need more time than others to demonstrate the desired levels of competence. Holding schools accountable for student performance on the first attempt seems to run counter to the belief that students learn at different rates, and the role of the school (particularly with respect to the MAAP) is to help as many students as possible to eventually pass. (This would not have been a major issue for the FLE, because almost everybody passed on the first attempt. We do not anticipate that occurring on the MAAP.)

Recommendation 61:
The Mississippi Department of Education may wish to consider changing the attempt after which schools are held accountable for a specified proportion of students passing the test. We believe it would be preferable to use the cumulative proportion who have passed at the end of grade 10, 11, or at the end of grade 12. Another possible consideration would be to use a standard that included the percent passing at two different grades (e.g., 75% pass rate in each test by the end of 10th grade and 85% pass rate after 11th grade.[28]

It has been proposed that the MAAP be administered for the first time at grade 10. The MAAP will have very high stakes for students who cannot graduate without performing satisfactorily on the tests. Results from these tests will also have high stakes for the high school education community. However, the lower elementary and middle schools will be concerned primarily about preparing students to do well on the norm-referenced tests included in the school accreditation system. In the past, norm-referenced tests administered in the lower grades have been weighted three times as heavily as the FLE (i.e., there were three performance standards for the NRT and only one for the FLE). With the change in 1994-1995 to administering norm-referenced tests in grades 4-9, the norm-referenced test results could be thought of as

counting six times as much as the FLE (or the new MAAP when it starts counting for accreditation). Indicators used in the accreditation system will drive at all levels what teachers teach and what students learn. The Department needs to be concerned that whatever elementary and middle schools are held accountable for teaching and students for learning be similar to, and/or provide a solid foundation for, what is measured by the new exams. Otherwise, elementary and middle schools could be preparing students very well for the tests administered at the lower grades and be rewarded for doing a good job, but find that students entering 10th grade are not well prepared to pass the exit exam, and some may not graduate as a result. If this should occur, both students and high schools will pay an extremely high price because of the lack of alignment.

Recommendation 62:
Performance standards established for the accreditation of school districts should be appropriately aligned and weighted. MDE should study carefully the alignment and weighting of performance standards used across the elementary, middle, and high school grades.

The Performance Standards for accreditation are defined in Bulletin 171, Revised 1994, pages 29-31. The "annual minimum value" (AMV) for the criterion-referenced tests (presumably including FLE and the end-of-course exam in Algebra) is set at a point that is one-half of an individual standard deviation below the mean score for all students tested, but this AMV is not allowed to fall below 70% correct on any of the criterion-referenced tests. Although it is possible to build tests intending to have specifications resulting in similar mean percentage correct scores, there is no indication that such a specification will be given much priority when a new exit exam or when new end-of-course exams are built. Furthermore, good test construction should not pay attention to such a requirement. Again, since the FLE has a ceiling effect for all three tests (even on the first attempt), this requirement for a minimum 70% or 80% correct regardless of what is measured or how it is measured has not been a problem on the FLE. With the introduction of the (assumed) more rigorous MAAP however, MDE may find that maintaining the same minimum average percent correct across the three test areas is just not sensible. There are simply no a priori reasons why the expectations for the students' performances should be the same across three different outcomes in different curricular areas. (Note: if the MDE should change its performance standards to include cumulative pass rates—as we recommend above—our concern would shift to how the passing standards are established for each test area. Again, setting the same percentage correct raw score for all three tests areas is not likely to be appropriate.)

Recommendation 63:
Remove from the revised performance standards any reference to minimum percent correct.

We infer from reading Bulletin 171 that the End-of-Course Exams will be included as performance standards in the Accreditation System as they are implemented. As such, these exams will have high stakes for schools but not necessarily for students. The Department needs to think carefully about the implications of this. Is this setting up a potential conflict between how the districts and how the parents/students would like to see resources allocated? Are districts going to be allowed to use the results of the end-of-course tests at the individual level (e.g., by allowing students to count their scores on such tests as a part of their course grades)? If some schools do allow the results to be used at the individual student level and other schools do not, how will this differential impact on students and their motivation to take the tests seriously impact the fairness of the accountability system?

As more of these end-of-course exams are added, the weight of the exit exams in the accreditation system will be reduced even further. It is conceivable that a student may have passed an end-of-course exam at grade 9 at a level demonstrating performance superior to that required to pass the more generic exit exam in the broad curricular area, but still be required to sit in 10th grade for the entire exit exam, including the generic test in a subject area already tested at grade 9.

Recommendation 64:
As new end-of-course exams are brought on line, MDE should study the extent to which these exams measure knowledge and/or skills included in the exit exam, review and perhaps expand the purpose for administering these tests, and evaluate again the proper weights such tests should have as compared to other performance standards included in the accreditation system (especially as compared to the tests students are required to pass in order to graduate).

Finally, we wish to make a comment about the constructed response sections on the norm-referenced tests administered in grades 4-9. As we understand the current plan, results on these sections are not to be counted in an accreditation system. This may not be wise. If they are not counted, the reforms that the MDE seeks may not be realized.

Recommendation 65:
Review the performance standards for the NRTs in grades 4-9 to determine whether or not it is possible and advisable to incorporate results from the constructed response sections as indicators in the accreditation system.

Section V:

Conclusions

WE HAVE DISCUSSED A NUMBER OF issues, offered a number of recommendations, and presented illustrative tasks to be performed with suggested completion dates for a state-mandated high school graduation test. We have also made some recommendations with respect to using this test and others for accreditation purposes.

It is clearly possible to develop a well-designed high school graduation test that meets curriculum, psychometric, educational, legal, administrative, and resource requirements. However, as this document has undoubtedly made clear, the task is not easy. For the task to be done well, a variety of steps need to be completed. For these steps to be completed, adequate funding needs to be made available.

While our recommendations will not all be repeated here, we point out below some of the aspects that have been considered in the report.

- It is legally inappropriate to hold students accountable for passing an assessment that covers material that they have not been taught. This makes using a high stakes graduation assessment to drive curricular change somewhat troublesome. One can use the announcement of an upcoming assessment to drive curricular change. This, of course, requires that there be considerable time between the announcement of the assessment and its implementation.

- Multiple-choice items can measure higher-order thinking skills and procedures. Performance assessments may not offer high enough psychometric qualities to be used for high stakes assessments. Mississippi certainly should not use performance assessments to measure those competencies that can be assessed with multiple-choice items.

- It is unlikely that any "off-the-shelf" test would be an acceptable high school exit test for the students of Mississippi.

- Requiring any national norm-referencing component of the exit exam would complicate the task of maintaining curricular validity for the test.

- There must be close articulation among the various assessment programs. They should not work at cross purposes, and if they are serving the same purposes, perhaps less assessment is needed.

- The use of the various tests in a performance-based accreditation model requires careful thought regarding how to set the performance level and what metric to use in setting the level (e.g., average performance or percentage of students above some cut score). ◆

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education., (1985). Standards for Educational and Psychological Testing. Washington: D.C.: American Psychological Association.

Debra P. v. Turlington, 564 F. Supp. 177 (M.D. Fla. 1983), aff'd, 730 F.2d 1405 (11th Cir. 1984).

Mehrens, W.A. (1993). Issues and Recommendations Regarding Implementation of High School Graduation Tests. North Central Regional Educational Laboratory. Oak Brook, IL.

Wainer, H. & Thissen, D. (1993). Combining Multiple-Choice and Constructed-Response Test Scores: Toward a Marxist Theory of Test Construction. Applied Measurement in Education. 6 (2), 103-118.

Wolcott, W. with Hoffman, A. (1994, June). Moving Forward: A Survey of Alternative Assessment in Literature and Current Practice. Office of Instructional Resources, University of Florida.

END NOTES

[1]Note that all recommendations follow rather than precede the relevant discussion.

[2]One of the non-panel reviewers has pointed out that restricting the assessments to reading, mathematics and written communication may not fit totally with the Superintendent's exhortation that the assessment be designed in such a way that education will have no alternative but to change dramatically. Further, the major stakeholders external to the public education arena may have expectations in additional areas. While we recognize this, it seems prudent to begin at a fairly modest level. As our recommendation points out, additional areas may be added at a later date.

[3]It is unclear to us how the meshing of English/Language Arts and Reading into one curriculum will impact the decision to have separate exit exams in reading and written communication although one of the non-panel reviewers of the previous draft posited that this would not likely be a problem.

[4]One non-panel reviewer of the previous draft eloquently articulated the need to focus on the "commitment to dramatic educational change." Another of the non-panel reviewers suggested much the same thing. As that reviewer stated, "I would suggest that we should be looking at criteria for assessment that identify those skills, knowledge, attitudes, and applications of knowledge that students should have. Whether those things are currently taught is a distractor. Assume they are not! Now the challenge becomes the creation of an assessment whose curricular and instructional validity will be established over time. The impact is that such a test and its initial results must be used as baseline for school improvement, and not for accountability." We would like to make clear that our stance is not against curricular change or high standards. However, there are legal and moral issues at stake when one implements a high-stakes test over untaught material and deprives some students of a diploma. As we point out in point 1 in the conclusions section, the change must precede the implementation of the test. An announcement of a high-stakes test over new and demanding content to be implemented in the future might legitimately serve as a catalyst for curricular change. Using the test initially as a catalyst for school improvement is acceptable. What is not acceptable is to use it for student accountability prior to establishing that students have had the opportunity to learn the material.

[5]One non-panel reviewer suggested that what should be stressed is the joint nature of the responsibility. Both the state and the local districts are responsible. We concur, but our report is primarily focused on the state responsibilities.

[6]If future documents are more like what we are describing as frameworks, they should be referred to as

such. If they are more like the current documents, we believe they should continue to be called curriculum structures.

[7]One non-panel reviewer stressed the importance of this recommendation and pointed out "that any effort to save funds by not carefully monitoring contractor work on a daily basis portends real problems. ...the analogy to daily site visits when building a new home would not be an exaggeration."

[8]Any such items would need to be included as a part of subsequent item tryout and pilot studies.

[9]One non-panel reviewer self described as "a strong advocate of performance assessment" commented as follows with respect to recommendations 16 and 17. "I concur with your conclusion in Recommendation 16 about performance assessment and high stakes individual test scores. Our experience has been that performance assessments can produce valid, reliable scores at the grade and content level by school. Use for high stakes individual graduation requirement testing is problematic at the present time, particularly because of the psychometric demands of number of independent measures and testing time necessary to obtain such measures. If Mississippi is, however, going to use the assessments for school improvement purposes, performance assessment offers a richness and opportunity for curricular integration that should be seriously considered." We would concur. It is in the context of using performance assessment in high-stakes exit examinations that our psychometric cautions should be heeded most carefully. Of course, even for school improvement purposes, one desires accurate assessment and – other things being equal – low costs are preferable to high costs.

[10]See our related discussion of timing in the Legal Issues section on p. 40.

[11]A partial compensatory model may also have implications for other technical considerations such as reliability, scaling, and equating. One can not look at the cut score process in isolation.

[12]One non-panel reviewer of the previous draft questioned whether the majority should be teachers.

[13]The composition of this committee is discussed later in the report.

[14]The training of this committee and the running of the item sensitivity reviews could be made a part of the test development contract. However, thought should be given to whether this overview could be perceived as a conflict of interest for the contractor in as much as the contractor wrote the original items and may be perceived as having a vested interest in keeping the items.

[15]This is not technically a reliability estimate, it is an estimate of the consistency of scoring. However, it often is referred to as interrater reliability.

[16]We do not have full technical details concerning how the scaled scores for the FLE were derived. We assume the scores are linear transformations from either logits or raw scores setting the scaled score of 233 at the cut score (70% correct) and the scaled score of 211 at 60% correct.

[17]Assessments for out-of-school adults should probably be under the control of the LEAs and be done at those same sites. Otherwise, the security problems increase immensely.

[18]It is possible that the child has not had an initial opportunity to learn the required content and skills. Thus, remediation may not be precisely the correct term. Also, we should point out that appropriate early assessments to identify weaknesses coupled with developmental instructional efforts should make remediation after the test less necessary than if such early detection and intervention efforts do not occur.

[19]Ideally the additional instruction should be provided in a manner and at times that do not take away opportunities to learn in other domains.

[20]One non-panel reviewer commented that "a strategy worthy of consideration might be to have a notification which must be signed by both the student and parent that clearly identifies that both parent and student understand that the passage of the test is a graduation requirement."

[21]Actually, if the first field testing of the items takes place in the fall of 1996, there is no reason not to notify them as of that date.

[22]See our final section on accreditation regarding use of the MAAP for that purpose.

[23]An example of what may be either incomplete documentation or inadequate communication is on the source and date of some of the documents we were given to review. Ideally, every document should be dated and the source of the document should be evident.

[24]A disadvantage of this approach is that the members of the bias committee may not feel as free to make comments about which items they believe are biased if they can not do so anonymously.

[25]One member of the panel believes the technical advisory committee should recommend the cut score. The remainder of the panel believes that it is important that the cut score be recommended by Mississippi educators. Either way, we all agree that the commonly used expression "setting" the cut score is a slight misnomer. Really, the committee makes a recommendation which is forwarded to other groups. The actual setting of the cut score is done by a governmental agency that has the power to make such a determination. That agency uses the information from the standard setting committee.

[26]One member of the panel strongly recommends using a single contractor (with three or four phases of the contract) for test development, field testing and test implementation. For continuity reasons this would be preferable if one can find a contractor that indeed is best at all these different components.

[27]One could also consider progress toward a goal as a criterion in the accreditation process.

[28]This recommendation assumes that the MDE will maintain cumulative pass rate information.

Ordering Information

1. Complete order form and mail with check or purchase order to NEFEC/SERVE, Route 1, Box 8500, 3841 Reid Street, Palatka, FL 32177 (904)329-3847. Make check or purchase order out to NEFEC/SERVE (Federal ID#: 59-6000-821).

2. Discounts are available for most SERVE products when you purchase units of fifty or more. For titles marked with an asterisk (*), quantities of 50-99 cost $6.00 each; orders of 100 or more are $5.00 each.

3. If you are requesting an invoice, your order must include a purchase order number.

4. We ship by U.S. Mail and United Parcel Service. Calculate your shipping charges from the table below. Shipping charges will be higher for special orders and shipping outside the continental U.S. Please allow two weeks for delivery from the time we receive the order in our office. If you require special shipping arrangements, let us know. In most cases, we can accommodate your needs.

5. For more information regarding SERVE's products and ordering procedures, please call 800-352-6001.

Shipping & Handling Charges

Up to $30.00	$ 2.50
$30.01 to $60.00	$ 5.00
$60.01 to $100.00	$ 7.50
$100.01 to $150.00	$ 10.00
$150.01 to $200.00	$ 12.50
$200.01 & above	call for price

Publications Listing

Description	*Item #*	*Price*
Appreciating Differences: Teaching and Learning in a Culturally Diverse Classroom	**HTADI**	**$7.00***
Assessment in Early Childhood Education: Status of the Issue	ECESI	FREE
Children Exposed to Drugs: Meeting Their Needs	**HTSEC**	**$7.00**
Children Exposed to Drugs: What Policymakers Can Do	PBCED	$1.00
Comprehensive School Improvement	**HTCSI**	**$7.00***
Designing Teacher Evaluation Systems that Support Professional Growth	RDTES	$7.00
Future Plans Planning Guide	**FPPLG**	**$7.00**
How to Assess Student Performance in Science: Going Beyond Multiple-Choice Tests	RDSPS	$7.00
Interagency Collaboration: Improving the Delivery of Services to Children & Families	**HTICO**	**$7.00***
Issues to Consider in Moving Beyond a Minimal Competency High School Graduation Test	RDMCT	FREE
Learning by Serving: A Compendium of Ideas for Service Learning	**HTLBS**	**$7.00**
The Need for Improved Mathematics and Science Education	PBIMS	$1.00
A New Framework for School Accountability Systems	**RDFRA**	**$3.00**
Overcoming Barriers to School Reform in the Southeast	RDBAR	$3.00
Promising Programs in Native Education	**LNPNE**	**$5.00**
Reducing Baby Bottle Tooth Decay	RBBTD	FREE
Reducing School Violence	**HTRVR**	**$7.00***
Reengineering High Schools for Student Success	HTRHS	$7.00
A Resource Book on Transitions	**TRANRB**	**FREE**
Rural Schools on the Road to Reform	EDRUR	$7.00*
Schools for the 21st Century: New Roles for Teachers and Principals	**HTSTC**	**$7.00***
Sharing Success in the Southeast: Mathematics and Science Education	SSMSI	FREE
Sharing Success in the Southeast: Preschool-to-School Transition	**SSPST**	**FREE**
Sharing Success: Promising Service-Learning Programs	SSPSL	FREE
Southern Crossroads: A Demographic Look at the Southeast	**SRSCR**	**$7.00***
Special Offer #1--School Improvement	P01	$25.00
Special Offer #2--Future Plans Video & Discussion & Planning Guides	**P02**	**$25.00**
Special Offer #3--Technology, Mathematics, and Science	P03	$25.00
Special Offer #4--Southern Crossroads	**P04**	**$25.00**
Supporting Family Involvement in Early Childhood Education: A Guide for Business	SRSFI	$5.00
Surveying the Landscape of State Educational Assessment Programs	**EDSYL**	**$7.00**
Teachers of the Year Speak Out: Key Issues in Teacher Professionalization	PBPIT	$1.00
Using Technology to Improve Teaching and Learning	**HTTEC**	**$7.00***
Youth Apprenticeship: A School-to-Work Transition Program	HTYAP	$7.00

Videotape Listing

Drug-Free Schools: A Generation of Hope	**VTDFS**	**$19.95**
Future Plans (Videotape) and Discussion Guide	FPPAK	$19.95
Passages: Providing Continuity from Preschool to School	**VTPST**	**$19.95**
School Improvement: Journey Toward Change	VTCSI	$19.95
Southern Crossroads: A Demographic look at the Southeast	**VTSCR**	**$19.95**
Successful Mathematics and Science Practices: General Audiences	EVTMS3	$19.95
Successful Mathematics and Science Practices: Policymakers	**VTMS6**	**$19.95**
Successful Mathematics and Science Practices: Teachers/Practitioners	VTMS9	$19.95

Order Form

Name: _____

Title: _____

Address: ❑ home ❑ work _____

City: _____ State: _____ Zip: _____

Phone: ❑ home ❑ work ( ) _____ Fax: ❑ home ❑ work ( ) _____

Quantity	Description	Item No.	Unit Price	Total

Subtotal	
Non-exempt Florida residents add 6% sales tax:	
S & H*	
Total	

Mail to:
NEFEC/SERVE
Route 1
Box 8500
3841 Reid Street
Palatka, FL 32177

Fla. Tax Exemption #:

Please photocopy this form for future orders.